AN INVESTIGATION OF THE SPECIFICITY OF GUINEA PIG LIVER TRANSGLUTAMINASE TOWARDS PROTEIN SUBSTRATES x

BY

PETER JOHN COUSSONS

(A thesis submitted for the degree of Doctor of Philosophy *).

x This work was jointly funded by the S.E.R.C. [Science and Engineering Research Council] and Celltech Ltd., as a C.A.S.E. [Co-operative Awards in Science and Engineering] studentship

* The information presented is based on work conducted between September 1987 and September 1990 in the Department of Biological and Molecular Sciences. University of Stirling, Stirling, SCOTLAND.

TEXT CUT OFF IN THE ORIGINAL

DEDICATION AND ACKNOWLEDGEMENTS

This thesis is dedicated to my father, Harry Coussons. His encouragement and financial assistance over my years in Stirling, made this post-graduate period of study possible.

For their involvement in collaborative efforts, supply of materials, and use of equipment, thanks to ; Dr. Christopher Johnson, Dr. Lindsay Sawyer, Dr. Bryan Smith, Ms. Sharon Kelly, Dr. L. Fothergill-Gilmore, Dr. J. Fothergill, Dr. A. Coulson, Dr. P. Cleat, Dr. L. Stevens, Dr. M. North, Dr. P. Brophy.

I should also like to thank a number of friends, family and colleagues, whose support and influence must have contributed in a less direct, though no less important manner to the style of research and completion of this project. Particular thanks to my mother, Rita Coussons, and my grandmother, Edith Coussons. Thanks also to ; Dr Cristina Fanutti, Dr Robert Briner, Dr Clive Fox, Patricia Briner, and Shirley Anderson.

Finally I would like to thank my supervisor Dr Nicholas Price for his interest, enthusiasm, caution, and many helpful discussions, over the last four years.

QUOTATION

The following quotation concerns Dr Richard Feynman's thoughts about scientific conferences. It describes a feeling which I have come to know well over this period of study.

"At each meeting it always seems to me that very little progress is made. Nevertheless, if you look over any reasonable length of time, a few years say, you find a fantastic progress and it is hard to understand how that can happen at the same time that nothing is happening at any one moment (Zenos paradox).
I think it is something like the way clouds change in the sky - they gradually fade out here and build up there and if you look later it is different. What happens in a meeting is that certain things which were brought up at the last meeting as suggestions come into focus as realities. They drag along with them other things about which a great deal is discussed and which will become realities in focus at the next meeting".

[R.P. Feynman (1961) summary of the Aix-en-Provence
Conference on particle physics].

It is my hope that some of the ideas presented herein will become some of the "realities in focus" for this small corner in the world of enzymology.

SUMMARY

The specificity of guinea pig liver transglutaminase was investigated by the determination of modification sites within polypeptides and proteins of known sequence and (in some cases) folded structure. It was shown that some globular proteins have substrate properties for transglutaminase in conformations which resemble their native states. Novel substrate sites for transglutaminase were determined within the following proteins:

1) bovine ß-lactoglobulin.
2) the $His_{388} \longrightarrow Gln_{388}$ mutant form of yeast phosphoglycerate kinase.
3) bovine ß-casein.
4) porcine pepsin.

Despite the high exposure of many glutaminyl residues within these proteins only a small fraction of these residues were observed to be reactive towards transglutaminase. This is taken to indicate that features such as the chemical nature of the amino acid side chains in the local vicinity of unreactive glutaminyl residues strongly determine the specificity of transglutaminase.

When structural models were available for substrates of transglutaminase, the local secondary structure associated with substrate sites could be assessed. When no such models were available computer based methods were used to predict the local secondary structures associated with these sites. This approach allows substrate sites to be classified according to their local conformational preference into conformationally flexible (type A substrates) and more conformationally restricted (type B substrates).

Since diverse amino acid sequences are observed to surround the reactive glutaminyl residues of many of the non-physiological substrates of transglutaminase, it was assumed that the glutaminyl residues within these sequences were probably reactive due to having;

1) a favourable stereochemistry during modification by transglutaminase
2) a lack of inhibitory features.

In order to determine why some exposed glutaminyl residues were reactive and others not, it was necessary to find features which were present in unreactive sequences but absent in reactive ones. Through the use of this approach an "anti-consensus sequence" motif was identified. This was based on the observation that exposed glutaminyl residues which were unreactive towards transglutaminase often have charged residues within their surrounding sequences. The distribution of allowed/disallowed residues within substrate sequences, together with what is known concerning the conformational preference of transglutaminase for its substrates was built into a preliminary set of "rules". These rules may provide a basis for understanding the observed specificity of transglutaminase. The application of these rules to a number of model systems has resulted in the correct prediction of both reactive and unreactive glutaminyl side chains within a number of proteins.

The demonstration of the substrate properties of the $His_{388} \longrightarrow Gln_{388}$ mutant of phosphoglycerate kinase illustrates the feasibility of introducing a novel substrate site for transglutaminase into a protein using recombinant DNA technology.

TABLE OF CONTENTS

CHAPTER 2: MATERIALS AND METHODS

CHAPTER THREE: PURIFICATION OF TRANSGLUTAMINASE FROM GUINEA PIG LIVER

CHAPTER FOUR: PRELIMINARY SCREENING OF PROTEINS AS SUBSTRATES

CHAPTER FIVE: AN INVESTIGATION OF THE SUBSTRATE PROPERTIES OF β-LACTOGLOBULIN (BOVINE) FOR GUINEA PIG LIVER TRANSGLUTAMINASE

CHAPTER SIX: AN INVESTIGATION OF THE SUBSTRATE PROPERTIES OF WILD TYPE AND MUTANT PHOSPHOGLYCERATE KINASES (YEAST) TOWARDS TRANSGLUTAMINASE

CHAPTER SEVEN: AN INVESTIGATION INTO THE SUBSTRATE PROPERTIES OF BOVINE β-CASEIN TOWARDS TRANSGLUTAMINASE

CHAPTER EIGHT: LESS WELL CHARACTERIZED SUBSTRATES OF GUINEA PIG LIVER TRANSGLUTAMINASE

CHAPTER THIRTEEN: CONCLUSIONS AND FUTURE WORK

APPLICATION OF THE CHARGE RULES TO FURTHER PROTEIN SUBSTRATES OF TRANSGLUTAMINASE (APPENDIX I)

APPENDIX II ASSAYS

APPENDIX III PRIMARY SEQUENCES OF PROTEINS

ABBREVIATIONS

ADP	:-Adenosine diphosphate
AMP	:-Adenosinemonophosphate
ATP	:-Adenosine triphosphate
Boc-	:-N-tertbutyloxycarbonyl group
BSA	:-Bovine serum albumin
C-	:-C-terminal carboxyl group of a peptide or protein
ß-CAS	:-bovine ß-casein
CBZ-	:-Benzyloxycarbonyl group
c.d.	:-Circular dichroism
cDNA	:-Complementary Deoxyribonucleic acid
D-	:-Dextrorotatory
DAP	:-distributer array processor
DNAase	:-Deoxyribonuclease
DNSC	:-dansylcadaverine
DTT	:-Dithreothreitol
EDTA	:-Ethylenediamine tetraacetic acid
FAB	:-Fast atom bombardment
FIB	:-fibrin
FIBRON	;-fibronectin
GDP	:-Guanosine diphosphate
Glc-	:-Glucosyl-
GMP	:-Guanosine monophosphate
GTP	:-Guanosine triphosphate
Kcat	:-Catalytic rate constant
kDa	:-Kilodalton
Km	:-Michaelis constant
Hb-	:-Haemoglobin
HLA-	:-Major histocompatibility antigen
HPLC	:-High pressure liquid chromatography
IgG	:-Immunoglobulin G
Ins	:-Insulin
L-	:-Laevorotatory
α 2-MAC	:-α_2-macroglobulin
MES	:-(2-[N-Morpholino]ethane] sulphonic acid)
mRNA	:-Messenger ribonucleic acid
N-	:-N-terminal amino group of a peptide or porotein
NAD	:-Nicotinamide adenosine dinucleotide
NADP+	:-Nicotinamide adenosine dinucleotide phosphate (oxidized)
NADPH	:-Nicotinamide adenosine dinucleotide phosphate (reduced)
nmr	:-nuclear magnetic resonance spectroscopy
ODC	:-Ornithine decarboxylase
PAGE	:-Polyacryamide gel electrophoresis
PCA-	:-Pyrrolidone carboxylic acid
pE-	:-pyroglutamic acid group
SDS	;-Sodium dodecyl sulphate
TRIS	:-(Tris[hydroxymethyl]amino-methane)
PGK	:-phosphoglycerate kinase
PGM	:-Phosphoglycerate mutase
α 2-P.I.	:-α_2-plasmin inhibitor
PTH-	:-Phenylthiohydantoin derivative
RHOD	:-rhodopsin

SV-IV	:-Seminal vesicle protein (Type IV)
TCA	:-Trichloroacetic acid
TEMED	:-N,N,N',N'-tetramethyletylenediamine
TFE	:-Trifluoroethanol
TPCK	:-N-p-toluene sulphonyl-L-phenylalanine chloromethylketone
TLCK	:-N-p-toluene sulphonyl-L-lysine chloromethylketone
u.v.	:-ultraviolet
VITRO	:-Vitronectin
Xaa	:-Unidentified PTH-derivative of an amino acid

Table 1.
Abbreviations of amino acids

Amino acid	Three letter abbreviation	One letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or Aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or Glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

The designation for some post-translationally modified proteins discussed in this thesis include X = Sulphonic acid derivative of a cysteine residue, and Xaa = unknown amino acid (or derivative).

CHAPTER 1: INTRODUCTION

1.1. THE REACTIONS OF TRANSGLUTAMINASE(S)

The term transglutaminase was first used by Heinrich Waelsch and coworkers (Clarke et al., 1957) to describe a calcium dependent transamidating activity present in a number of tissues, which could be measured by the covalent incorporation of amines (eg. histamine and putrescine) into a soluble protein fraction of mammalian liver homogenates. Transglutaminases are now known to catalyse a Ca^{2+} dependent aminolysis of glutaminyl side-chains via an acyl-transfer mechanism in which the γ-carboxamide group acts as an acyl donor and suitable primary amines, water, or the ε-amino group of a peptidyl lysine side chain may act as acyl acceptors. These enzymes are sometimes termed protein-glutamine: amine-glutamyl-transferases, and appear to be involved with diverse biological functions (see section 1.10.). **This thesis is primarily concerned with the specificity and general biochemistry of the cytoplasmic transglutaminase of guinea pig liver (EC 2.3.2.13.). Thus the term "transglutaminase" must be taken to refer to the enzyme from this source unless otherwise stated.**

1.1.1. THE PURPOSE OF THE INTRODUCTION

The relationship between structure and function is a central paradigm of molecular biology. Thus it is highly unlikely that the physiological role of an enzyme such as transglutaminase (see section 1.11.) may be appreciated in the absence of an understanding of the basis of its specificity. The introduction to this thesis is intended to give the reader sufficient background to the biochemistry of the transglutaminases so

that the discussion of the specificity of the enzyme may be viewed in relationship to its physiological roles *in vivo*.

The phylogenetic and tissue distribution of transglutaminases are considered in sections 1.1 and 1.2. The genes coding for a number of transglutaminases have been cloned and sequenced from a number of these tissues. The similarities in the structure of transglutaminases (see section 1.3) and the cysteine proteinases suggest that they may be distantly related (see section 1.4). This similarity extends to a common catalytic mechanism (see section 1.5.). Section 1.6. is designed as a general guide to the handling of transglutaminase and describes some of the general features which will be of interest to the protein chemist.

Reviews concerning the general chemistry and physiological role(s) of transglutaminases have been published and are available [Lorand and Conrad (1984) ; Zappia et al., (1988)]. More specialized subject areas are reviewed elsewhere and these include; "The structure of transglutaminases": (Ichinose et al., 1990); "Mechanism and basis for specificity of transglutaminase catalysed isopeptide bond formation" (Folk, 1983). This review is specifically concerned with the activity of the liver transglutaminase and factor XIIIa towards small synthetic amide and amine substrates. To date (October, 1992) there has been no published review concerning the specificity of transglutaminases towards protein and polypeptide substrates and this subject is therefore considered in detail in sections 1.7. and 1.8.

1.2. DISTRIBUTION OF TRANSGLUTAMINASE(S)

1.2.1. PHYLOGENETIC DISTRIBUTION

Calcium dependent transglutaminases are widely distributed in tissues and body fluids (see section 1.2.2.) of most animals tested eg. sea urchin, (Cariello et al., 1984), human (Lorand et al., 1981), and rat (Wong et al., 1990). Recently a novel form has been detected in the adult worms of the filarial parasite *Brugia malayi* (Mehta et al. 1990). This 22 kDa protein was only found in female worms.

Non-calcium dependent transglutaminases have been recently discovered in plant tissues (Margosiak et al., 1990), and bacteria (Ando et al., 1989). The amino acid sequences of these two forms of transglutaminase are not known and hence their evolutionary relationship to the calcium dependent transglutaminases is enigmatic.

1.2.2. TISSUE DISTRIBUTION

Transglutaminase activity is widely distributed amongst mammalian tissues and is known to be due to the expression of a number of sequence related proteins (see section 1.3).

Some tissues appear to be constitutive in this activity while in others it is usually absent unless induced (see section 1.10.). The same form of transglutaminase may be found in different tissues eg. the intracellular 80 kDa form is found in both liver and lens tissue (Lorand et al., 1981). Other tissues in which activity has been detected include: hair follicles, (Harding and Rodgers, 1972), epidermis, (Simon and Green, 1985), blood

plasma, (Lorand et al., 1981) and the coagulating gland of the anterior prostate, (Porta et al., 1990). Multiple forms of transglutaminase have been detected in some tissues eg. two enzymically active bands, corresponding to transglutaminase were observed following separation of purified protein from guinea pig liver, using a non-denaturing electrophoretic technique [details not given] (Lorand and Conrad, 1984) (see section 1.3.5.).

1.3. THE STRUCTURE OF TRANSGLUTAMINASES

1.3.1. PRIMARY STRUCTURE

The complete amino acid sequence of guinea pig liver transglutaminase has been predicted from the sequence of cloned cDNA (originally complementary to its mRNA), (Ikura et al., 1988). The molecular weight was calculated to be 76620. The enzyme contains no carbohydrate, (Folk and Chung, 1973) despite the presence of six potential Asn-linked glycosylation sites and does not contain any disulphide bridges, (Folk and Cole, 1966). A hydropathy profile of the liver enzyme suggests that the active site is at the amino end of a region of high hydrophobicity, (Ikura et al., 1988). This may correspond to the hydrophobic pocket predicted by Folk and Gross (1971). No "E-F hand" structures (a common Ca^{2+} binding sequence motif), have been located. The sequences of a number of transglutaminases other than the liver enzyme have been determined. These include the factor XIIIa from human placenta (Takahashi et al., 1986), the human keratinocyte enzyme (Yamanishi et al., 1991), and others (see Ichinose et al., 1990). The sequences have been aligned using computer based techniques against the guinea pig liver sequence by Phillips et al., 1990. The evolutionary relatedness of these proteins is discussed in section 1.4..

1.3.2. SECONDARY STRUCTURE

Circular dichroism studies of transglutaminase suggest that liver transglutaminase has some stable secondary structure, (Folk and Chung, 1973). Calcium ions induce a conformational change which causes activation of the enzyme (see section 1.6.2.). This does not effect gross changes in size or shape of the enzyme molecule.

A comparison of the secondary structural predictions of different transglutaminases (see section 12.2.1.), suggests that these proteins have conserved secondary structure over large areas. Little α-helix is predicted and β-turn and β-sheet structures predominate, (Kim et al., 1990).

1.3.3. TERTIARY STRUCTURE

The placental protein has been crystallized, and an almost complete three dimensional data set collected, (Hilgenfeld et al., 1990). The crystals diffracted to a resolution limit of 2.7 Å. No three dimensional structural model has yet been presented. However electron microscopy studies of factor XIIIa indicates that the a_2 dimer consists of two globular particles of approximately 6 x 9 nm in size (Carrell et al., 1990). The extensive sequence similarity of the liver enzyme to factor XIIIa suggests that this protein probably also contains globular structure (see section 1.3.1.).

1.3.4. QUATERNARY STRUCTURE

The diversity of quaternary structure within the transglutaminase family of proteins is reviewed elsewhere, (Ichinose et al., 1990). Transglutaminases range in form from multi-subunit zymogens to single chain forms. These proteins may be composed of identical subunits or as large catalytic subunits and smaller regulatory subunits eg. plasma factor XIIIa. The cytoplasmic guinea pig liver transglutaminase appears to be a single stranded non zymogenic form.

1.3.5. POST-TRANSLATIONAL MODIFICATION OF TRANSGLUTAMINASE

A number of post-translational modifications have been identified within this protein family. The guinea pig liver enzyme undergoes amino-terminal processing, (Ikura et al., 1989). This involves removal of the N-terminal methionine and subsequent N-acetylation of the new N-terminal alanine residue. Factor XIIIa is activated following partial proteolysis by thrombin in the presence of fibrin (see section 1.4.), (Takagi and Doolittle, 1974). The keratinocyte transglutaminase is thought to be myristoylated within a membrane anchoring region located within 10 kD of the N- or C-terminus in a hydroxylamine-sensitive (possibly thioester) linkage, (Chakravarty and Rice, 1989). These and other presently unknown post-translational modifications may account for the multiple molecular forms of these enzymes within single tissues (Lorand and Conrad, 1984) (see also sections 1.2.2. and 1.11.4.4.).

1.4. EVOLUTION OF TRANSGLUTAMINASES

The amino acid sequences of the four known transglutaminases (see section 1.3.1.) are similar to each other and to protein Band 4.1, [a component of the membrane skeleton of red blood cells, (see section 1.10.4.)]. The similarity is particularly marked about the active site cysteine residue, (Phillips et al., 1990). Band 4.2 does not appear to have transglutaminase activity. The region of amino acid sequence within Band 4.2 which strongly resembles transglutaminase has an alanine residue in the position normally occupied by catalytically active cysteine residues in transglutaminases (Sung et al., 1990).

When identical amino acids are present at a given position in two of the four sequences, it is evident that the keratinocyte enzyme and factor XIIIa are most similar to each other. The same conclusion follows from percentages of amino acid identity after aligning the sequences two at a time. By contrast the tissue transglutaminase appears as closely related to the band 4.2 protein as to other transglutaminases. This is compatible with an initial divergence of the band 4.2 protein from the group, with subsequent separation of the tissue enzyme from this lineage. The catalytic subunits of factor XIIIa and the keratinocyte enzyme may have evolved from an ancient predecessor of the liver enzyme or a related protein.

The deduced evolutionary relationships among these enzymes suggest that N-terminal extensions seen in epidermal transglutaminases (Phillips et al., 1990) occurred for specialised functions after divergence of the tissue enzyme from a common lineage. This may also be the case in other transglutaminases. An N-terminal extension is proteolytically cleaved from factor XIIIa by plasmin during conversion of the zymogenic form of factor XIIIa (a_2b_2) to the enzymically active form ($2a'$ (Lorand et al., 1980). This

sequence is encoded in a separate exon to the rest of the a subunit. It is possible that the keratinocyte N-terminal extension and the activation peptide of factor XIIIa are derived from a common origin, followed by gene divergence (Ichinose et al., 1986).

The active site sequence of both the liver enzyme and the other transglutaminases shows considerable similarity to that of papain and other cysteine proteinases (Ikura et al., 1987). Both families of enzyme are known to have extended binding sites for protein substrates, and this relationship is discussed further in chapter 12.

1.5. THE CATALYTIC MECHANISM OF TRANSGLUTAMINASE(S)

The evidence relating to the mechanism of catalysis has been reviewed elsewhere, (Folk 1983). This model of the catalytic mode of action is illustrated diagrammatically in Fig 1.1..

The enzyme is thought to follow a modified double displacement mechanism. Initially the acyl donor (glutaminyl side chain or analogue) binds at the active site of the enzyme to form a covalent (though normally short-lived) thioester (Glu - Cys) intermediate with Cys_{275}. The enzyme is then thought to undergo a conformational change, concomitant with release of the ammonium leaving group. This would leave the enzyme-substrate complex available for hydrolysis by water to result in a deaminated (Gln → Glu) product. In the presence of an appropriate concentration of unbranched primary amine (eg. putrescine or a lysine side chain), a modified Gln product eg. ε-(γ-glutamyl)putrescine or ε-(γ-glutamyl)lysine is formed by nucleophilic attack on the thioester intermediate by the amine (the amine must effectively compete with water molecules at the active site). Following release of the product, the active site Cys_{275} is regenerated. The molecular structures of some of the

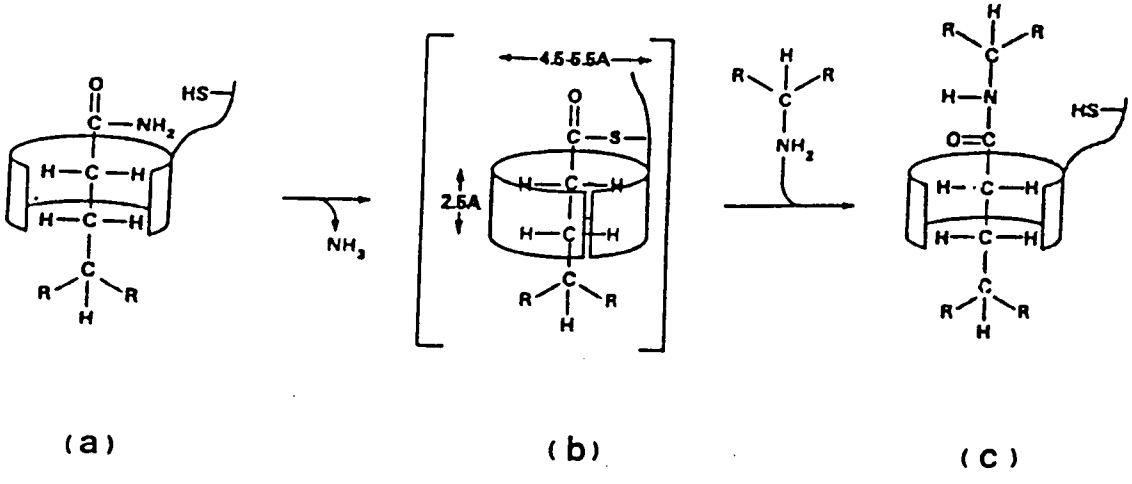
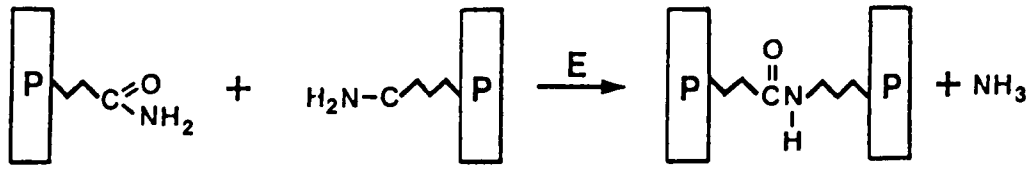


Fig 1.1. THE MECHANISM OF ACTION OF GUINEA PIG LIVER TRANSGLUTAMINASE.

The amide substrate (glutamine) is shown to bind at the active site of transglutaminase in diagram (a). The -SH group of the active site of the enzyme is then thought to attack the carboxamide group of the substrate to form a thioester intermediate (shown in b). There may be a conformational change at this point in the mechanism and this is depicted by the closing of the circular collar which surrounds the bound thioester in diagram (b). Finally the enzyme-thioester intermediate is thought to be attacked by an incoming amine (nucleophile) substrate (see Fig 1.3. for some typical structures) and the amine is shown to exchange for an ammonium ion resulting in:(i) the formation of a cross-linked product between the glutaminyl residue and the amine substrate (ii) the release of ammonia. This is shown in diagram (c) [see also Fig 1.2]. The diagrams are taken from Folk (1983).

REACTION (a)



REACTION (b)

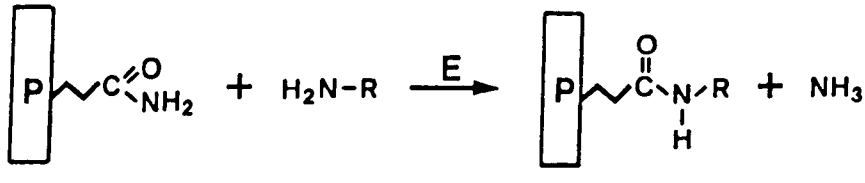


Fig 1.2. TWO TYPES OF BIOLOGICAL REACTIONS CATALYSED BY TGASE

Reaction (a) depicts the transglutaminase catalysed cross-linking of protein molecules via a ε-(γ-glutamyl)lysine bond between the two molecules. This may also provide the basis for stabilizing elements of intramolecular cytoskeleton within cells.
Reaction (b) shows the incorporation of small molecular weight amines into the γ-glutaminyl side-chains of proteins. This may be part of an important regulatory system in cells (see section 1.10.11.).

[Symbols; P = protein molecule, E = transglutaminase]

products of transpeptidation of a glutaminyl residue by transglutaminase, when either putrescine, or a lysine side chain are shown in Fig 1.2..

1.6. TRANSGLUTAMINASE AS A TOOL FOR PROTEIN CHEMISTRY

1.6.1. INTRODUCTION

The specificity (see section 1.7.), and mild reaction conditions required for the activity of transglutaminase (sections 1.6.2.- 1.6.6.) make it a useful tool for the protein chemist. Some general points concerning the activity of transglutaminases *in vitro* are considered below.

1.6.2. EFFECT OF CALCIUM ON TRANSGLUTAMINASE ACTIVITY

Tissue transglutaminase has been shown to be 90% activated at a concentration of 1 mM Ca^{2+}. (Clarke et al., 1959). The problems associated with the calculation of calcium dependency are considered by Cooke and Holbrook (1974).

1.6.3. STRUCTURAL FEATURES ASSOCIATED WITH AMINE SUBSTRATES OF TRANSGLUTAMINASE

All unbranched, uncharged, aliphatic amines tested (including hydroxylamine) have been shown to act as substrates for

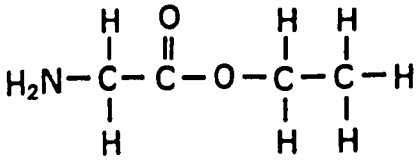
transglutaminase(s), (Folk, 1983 ; Clarke et al., 1959). Amines branched prior to the β-position and those with an adjacent negative charge (eg. aminobutyric acid) do not act as substrates, (Clarke et al., 1959). The order of reactivity at pH 7.5 towards some of the commonly used primary amines is : putrescine > glycinamide > histamine > spermine > cadaverine > methylamine > ethanolamine. Ie amines which resemble the side-chain of lysyl residues are the best substrates. The covalent structures of some of these substrates are shown in Fig 1.3..

In the absence of primary amine, transglutaminases catalyse deamination of glutaminyl residues (Folk, 1983). A concentration ranging between 1.0 to 5.0 mM amine is usually sufficient to inhibit this reaction.

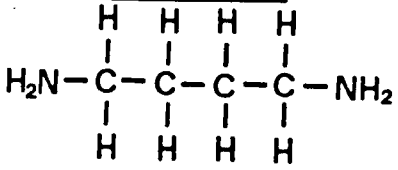
1.6.4. THE EFFECT OF TEMPERATURE ON TRANSGLUTAMINASE ACTIVITY

The curve for the dependence of the reaction with temperature, (using heat treated liver homogenate as a substrate) exhibits the bell shape characteristic of enzymic reactions, (Clarke et al., 1959). At pH 7.5 and 1 mM Ca^{2}+ the rate doubles for a 10 °C rise in temperature up to a maximum rate at between 30 and 37 °C and then falls off rapidly to less than 10% of the maximum at 55 °C. The enzyme is not thermostable since the half-time ($t_{1/2}$) of the deactivation ranges from 65 min at 40 °C to less than 30 s at 65 °C. This proposal is supported by the observation of Nury and Meunier, (1990) who have reported that the number of titratable (and hence exposed) cysteines in transglutaminase correlated directly with loss of activity when it was incubated at 55°C. Thus the inactivation of transglutaminase appears to be due to an initial unfolding of the polypeptide chain, followed by the aggregation of polypeptide chains. Inactivation via deamination or self cross-linking does not appear to be responsible for the observed loss of enzymic activity. Birckbichler

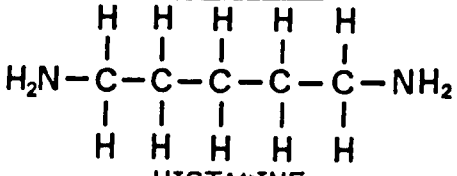
GLYCINE ETHYLESTER



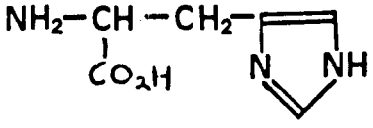
PUTRESCINE



CADAVERINE



HISTAMINE



PEPTIDYL LYSINE

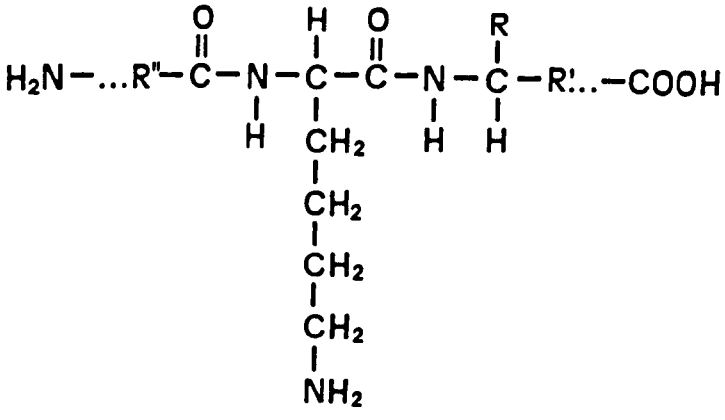


Fig 1.3. THE MOLECULAR STRUCTURES OF SOME AMINE SUBSTRATES OF TGASE

All of the commonly used amine substrates of transglutaminase shown above are highly soluble in water in the range from pH 6.0 to pH 9.0. Radioactive isomers of these compounds facilitate the calculation of the amount of amine bound to a protein by transglutaminase and assist in the location of the reactive sites within proteins during peptide sequencing. The structure of the commonly used fluorescent amine substrate, dansylcadaverine is shown in Fig 2.4.

et al., (1977) have reported that transglutaminase will catalyse incorporation of label into itself and cross-link itself into material of high molecular weight. This may have been due to partial denaturation following incubation at 37 °C for 1 hr.

1.6.5. THE EFFECT OF pH ON TRANSGLUTAMINASE ACTIVITY

Transglutaminase is active with a number of primary amine substrates over the range pH 5.0 to 9.0. The pH optimum with different amines was shown to vary between pH 7.8 and pH 8.3 (Clarke et al., 1959).

An ammonium salt such as tris(hydroxymethyl)aminomethane hydrochloride (Trizma) does not have substrate properties and hence is suitable as a buffer for *in vitro* work. Substances capable of combining with calcium, for example citrate, phosphate, pyrophosphate, and fluoride should be excluded from reaction mixtures.

1.6.6. THE EFFECT OF REDUCING AGENTS ON TRANSGLUTAMINASE ACTIVITY

An early report suggested that transglutaminase was dependent on the presence of sulphydryl activators (Block, 1954). This report led to the routine addition of reducing agents to the assay media of transglutaminase.

Brenner and Wold (1978) investigated the modification of globular proteins by transglutaminase in the absence of added reducing agent and observed a "greater degree of discrimination in the utilization of protein substrates" relative to mixtures in which DTT was included. Reducing agents such as DTT will cleave susceptible disulphide bonds within

proteins and this will tend to increase the exposure of reactive glutaminyl residues.

1.6.7. THE EFFECT OF NUCLEOTIDES ON TRANSGLUTAMINASE ACTIVITY

Transglutaminase activity is reduced by the binding of guanine nucleotides, such as GTP, (Bergamini et al., 1987). GTP binding appears to inhibit calcium binding and the associated conformational change at non-saturating Ca^{2+} concentrations, (Bergamini, 1988) [see section 1.5.].

Transglutaminase(s) have a low GTPase activity, (Lee et al., 1989). The inclusion of GTP, GDP, and GMP in reaction mixtures appears to inhibit the activity of both guinea pig liver transglutaminase, and factor XIIIa. It is possible that guanine nucleotides are regulators of transglutaminases *in vivo*. It may also be significant that the ε-(γ-glutamyl)lysine cross-link content of the intracellular proteins of the slime mould, *Physarum polycephalum* appear to be modulated *in vitro* by treatment with Mg^{2+}-ATP and Mg^{2+}-ATP plus calcium (Loewy and Matacic, 1981). The cross-links appear to be concentrated within the proteins of the cytoskeleton (see sections 1.10.3. and 1.11.3.), especially actin (see section 1.7.8.2.), fibronectin (see section 1.8), myosin, and two proteins of 52 kDa and 51 kDa.

1.7. THE DETERMINATION OF THE SPECIFICITY OF TRANSGLUTAMINASE TOWARDS GLUTAMINYL SIDE-CHAINS *IN VITRO*

1.7.1. INTRODUCTION

The specificity of transglutaminase has been investigated using a variety of approaches. It has been tested on :

1) denatured proteins which contain multiple glutaminyl sites, in order to examine the effect of different sequences surrounding glutaminyl residues on glutaminyl reactivity.

2) blocked synthetic dipeptides, to test the effect of neighbouring residues on glutaminyl reactivity.

3) larger synthetic peptides where the effect of single amino acid replacements in positions corresponding to a putative extended binding site could be assessed.

4) fully folded globular proteins to assess the local secondary structure associated with substrate sites which may be accommodated by transglutaminase.

The results of these investigations are reviewed in the following sections (1.7.2.-1.7.8.4.) under section headings which describe the general structure of substrate being considered. The specificity of transglutaminase towards small peptides is dealt with first, [This has been reviewed by Folk (1983)]. These substrates are expected to have highly flexible structures in solution. Secondly larger peptides and proteins, which have the potential to form secondary structure are discussed. Finally the specificity of transglutaminase towards globular proteins is considered. The apparent determinants of transglutaminase specificity are

discussed on the basis of the information collated from the preceding sections in section 1.7.9..

The substrate sites of factor XIIIa within proteins are compared against each other in section 1.8. A possible basis for the observed differences in specificity between factor XIIIa and the guinea pig liver enzyme is discussed in section 1.8.4.

NOTE: **Within the following sections reactive glutaminyl residues are considered to be located at position zero in a given sequence. Residues on the C- terminal side of the glutaminyl residues are designated positive positions (eg +1, +2 etc., and residues N- terminal to the glutaminyl residues are designated negative values relative to the glutaminyl residue in question.**

1.7.2. ALL PEPTIDYL-GLUTAMINYL RESIDUES ARE POTENTIAL SUBSTRATES FOR TRANSGLUTAMINASE

Using conditions in which the substrate properties of globular proteins were compared against their performic acid oxidized counterparts, Toda and Folk. (1969) showed that prolonged incubation (48 hr) with large quantities of transglutaminase is probably sufficient to cause modification of most glutaminyl sites within the following denatured proteins : lysozyme (chicken egg white), ribonuclease A (bovine pancreas), insulin (bovine pancreas) and haemoglobin (human). The non-denatured counterparts were generally far less reactive. In consideration of the variety of amino acid residues in close vicinity to glutamine residues within the protein substrates studied, the authors concluded :

"It seems that polypeptide conformation and not primary sequence renders certain glutamines totally resistant to transglutaminase".

While this observation is probably true, there are considerable differences in the reactivity of different glutaminyl residues within a polypeptide sequence and this is considered in sections 1.7.3. to 1.7.8.. The reactivity of glutaminyl residues within polypeptides and proteins towards transglutaminase appears to be strongly influenced by conformational factors which include:

1) local steric factors.
2) secondary structure,
3) tertiary structure,

Factors such as hydrophobic, electrostatic, and other chemical features associated with polypeptide sequences may also influence the reactivity of a glutaminyl site independently of their effect on polypeptide conformation.

1.7.3. SMALL ALIPHATIC AMIDES

All straight chain aliphatic amides tested (with the exception of formamide) have been shown to act as substrates for transglutaminase (Folk 1983). Amides branched at positions before the β-position are poor substrates. All transglutaminases tested to date are unreactive towards free glutamine. There is almost complete stereospecificity towards L-glutaminyl residues within synthetic peptide substrates.

1.7.4. SMALL SYNTHETIC BLOCKED DIPEPTIDES

The reactivity of the liver enzyme towards simple blocked dipeptides has been shown to be: CBZ-L-Gln-L-Val ethylester > peptides where the Val is replaced with another amino acid in the order: L-Tyrosine > L-Leucine > L-Alanine > L-Glycine > D-Valine > L-proline. Carboxyl-terminal glutaminyl residues are even poorer substrates (Neidle and Acs, 1961). The lengthening of the peptide chain between blocking group and the glutaminyl side-chain was shown to reduce the activity of peptides relative to CBZ-Gln-Gly eg. CBZ-Gln-Gly is ten times better as substrate than CBZ-Gly-Gln-Gly, and this in turn is ten times better than CBZ-Gly-Gly-Gln-Gly, (Gorman and Folk, 1980). Tetra and pentapeptides are less reactive relative to their carboxypeptide esters (Folk and Cole, 1965). Removal of the CBZ-blocking group abolished the substrate properties of these peptides. However larger peptides were found to retain activity on removal of the blocking group. The substrate properties of a number of larger unblocked synthetic peptides have been investigated, Folk (1983) [see section 1.7.5.].

1.7.5. LARGER UNBLOCKED SYNTHETIC PEPTIDES (4-15 residues)

1.7.5.1. INTRODUCTION

A number of peptides with sequences based on known labelling sites within protein substrates have been synthesized and tested as transglutaminase substrates. In virtually all cases the specificity of the liver enzyme appeared to be different from that of factor XIIIa. This is discussed further in section 1.8.4.

1.7.5.2. FIBRONECTIN PEPTIDE ANALOGUES

On the basis of half-saturations of the enzymes, the acceptor affinity of the peptides using dansylcadaverine as acyl acceptor were in the order shown in Fig 1.4. (Parameswaran et al., 1990).

Boc-Q-Q-I-V > Boc-A-Q-Q-I-V > p-E-A-Q-Q-I-V

Fig 1.4. SEQUENCES BASED ON FIBRONECTIN

[Boc = N-tertbutyloxycarbonyl , pE = pyroglutamic acid]

The acceptor affinity was in the reverse order for factor XIIIa. Cross-linking of both ß-Bp crystallin (see section 1.7.8.) and fibrin (see section 1.8.) by transglutaminase was effectively inhibited by Boc-Q-Q-I-V when used in conjunction with the liver enzyme. Whether only one or both glutaminyl residues were modified was not reported.

1.7.5.3. ß-CASEIN PEPTIDE ANALOGUES

The k_{cat}/Km (app) ratio has been used as a measure of the specificity of guinea pig liver transglutaminase and factor XIIIa for a variety of peptides modelled on the known substrate site of factor XIIIa in non-succinylated bovine ß-Casein (ie. Gln_{187}), shown in Fig 1.5. Peptides [1] and [2] represent parental sequences upon which a number of related peptides were designed. The effect of a number of single, and multiple

substitutions of residues by glycine along the sequence on the substrate properties of these peptides was investigated. The specificity of guinea pig liver transglutaminase towards these peptides appeared to be approximately twice that of the enzyme towards Gln_{167} in intact bovine ß-casein, and was different to that of factor XIIIa (Gorman and Folk, 1980) (see section 1.8.). Dansylcadaverine was used as the amine substrate.

PEPTIDE [1] NH_2-S-V-L-S-L-S-Q*-S-K-V-L-P-V-P-E-COOH

PEPTIDE [2] NH_2-L-G-L-G-Q*-G-K-V-L-G-$CONH_2$

Fig 1.5. MODEL PEPTIDES BASED ON BOVINE ß-CASEIN
The reactive glutaminyl residue is indicated with an asterix*

In peptide [1] most single amino acid replacements reduced substrate characteristics, (Gorman and Folk, 1981). Multiple replacements reduced them further.The simultaneous replacement of the residues at positions, +2, +4, and +6 with serine residues in peptide [2] (Gorman and Folk, 1984), had little effect on the substrate properties of the product. This may reflect the generally favourable chemical nature of small hydroxylated amino acids within transglutaminase substrates [noted by Yan and Wold (1984)]. Substitution of residues in the region C-terminal to the glutaminyl residue by glycine reduced substrate characteristics more than those N-terminal to this residue. This was taken to reflect a requirement for hydrophobic groups in this area.

The effect of succinylation on the substrate properties of a number of these peptides was tested. The effects were not explicable on the basis of increased steric blocking of reactive glutaminyl residues. Selective succinylation of the α-amino group had a variable effect depending on the peptide modified. Gorman and Folk (1984) have suggested that the reason for the high substrate effectiveness of bisuccinylated peptides could be due to the reduction of positive charge on the side chain of Lys_7.

1.7.5.4. FIBRIN PEPTIDE ANALOGUES

A synthetic tridecapeptide was designed on the basis of the known amino acid sequence of the substrate site for factor XIIIa within the C-terminal [Leu_{373} to Gly_{385}] region of the γ-chain of human fibrin (Chen and Doolittle, 1971). The sequence of this peptide is shown in Fig 1.6. The peptide is a poor substrate when compared to the parent protein or the succinylated γ-chain alone, when dansylcadaverine was used as the amine co-substrate (Gorman and Folk, 1980). Even lower specificities were observed for shorter peptides based on this region of fibrin.

1 L-T-I-G-E-G-Q-Q-H-H-L-G-G 13

Fig 1.6. MODEL PEPTIDE BASED ON FIBRIN

Sequence analysis of the dansylated peptide indicated that the majority of the dansylcadaverine was located in Gln_7 with little incorporation into

Gln_8 following modification by guinea pig liver transglutaminase. This indicates that Gln_7 is the more reactive of the two glutaminyl sites.

1.7.6. NATURALLY OCCURRING PEPTIDES (11-30 residues)

1.7.6.1.INTRODUCTION

A number of polypeptide hormones have been tested as model substrates for transglutaminase. These include: α and β-endorphin (Porta et al., 1988) [these correspond to the amino acid sequence from Tyr_{61} to Thr_{76} and the Tyr_{61} to Glu_{91} C-terminal fragments of the hormone β-lipotrophin (Li, 1964)], substance P, (Pucci et al. 1988), melittin, (Perez-Paya et al., 1991) and glucagon. (Folk and Cole, 1965). The labelling sites within these proteins have been aligned in Table 1.1..

1.7.6.2. PRIMARY AMINO ACID SEQUENCES OF SUBSTRATE SITES WITHIN POLYPEPTIDE HORMONES

The polypeptide substrates shown in Table 1.1. were all modified under similar conditions ie. saturating calcium concentration (5 mM) between pH 7.0 and pH 8.0 at temperatures in the range 25-37°C. A variety of primary amines were used as substrates (amine structure is not expected to strongly affect the specificity of transglutaminase for these glutaminyl residues, see section 1.6.3.).

Table 1.1. THE SEQUENCES OF SUBSTRATE SITES WITHIN POLYPEPTIDE HORMONES

PROTEIN	RESIDUE	AMINO ACID SEQUENCE *	RESIDUE
B-ENDORPHIN	4	F-M-T-S-(E)-K-S-Q-T-T-P-L-V-T-L	18
GLUCAGON	1	NH_2-H-S-Q-G-T-F-T-S-(D)-Y	10
GLUCAGON	17	T-L-(D)-S-R-R-A-Q-(D)-F-V-Q-W-L-(M)-COOH	31
SUBSTANCE P	1	NH_2-R-P-K-P-Q-Q-F-F-G-L-(M)-COOH	11
MELITTIN	18	S-W-I-K-R-K-R-Q-(O)-COOH	26

* The amino acid sequences are aligned against the reactive glutaminyl residue within the sequence shown. The alignment of some of the peptides on either side of the modified residue is limited in some cases by the termination of the chain at the amino (N-) or carboxy (COOH) ends.

[Positively charged residues = Lys and Arg Negatively charged residues = Asp and Glu].

The diversity of amino-acid sequences adjacent to the modification sites indicates that the specificity of transglutaminase for this type of substrate is quite broad. The positive charge located on the N-terminal side of the modified site in all cases, did not preclude modification at these sites. However in no cases are positively charged side-chains located nearer than nine residues to the C-terminal side of reactive glutaminyl residues. Negatively charged residues are observed to be distributed on both sides of reactive glutamine residues, and hence their presence does not appear to reduce substrate properties of the parent polypeptide in these cases.

1.7.6.3. SECONDARY STRUCTURAL FEATURES ASSOCIATED WITH POLYPEPTIDE HORMONES

Polypeptide hormones have characteristically flexible structures at neutral pH, (Blundell and Wood, 1982). The preferred conformation depends on whether the charge state of the molecule, its concentration and the solvent in which it is dissolved. Physicochemical techniques have indicated that even short peptides may have some secondary structure in solution. For example; Substance P has been predicted to have ß-sheet/ß-turn potential, (Williams and Weaver, 1990). ß-Lipotrophin, [Hollosi et al., (1977) ; Yang et al., (1977) ; Lichtarge, (1987), and melittin (Terwilliger and Eisenberg, 1982) appear to have some α-helical structure in aqueous solution.

Glucagon has been crystallized (Saski et al. 1975) and the three glutaminyl residues are located in regions which differ in local secondary structure. The first eight residues of the N-terminal region appear to have an extended conformation after which the peptide becomes increasingly more α-helical towards its C-terminus. It may be significant that Gln_{24} which is the only unreactive glutaminyl residue within glucagon appears to be sterically blocked by residues which neighbour it in its primary sequence. This is shown in Fig 1.7. The rigidity of local structure about Gln_{24} has been verified by a n.m.r. study (Blundell, 1979). Thus it is not clear what the overall conformational similarity of these polypeptides is at the active site of transglutaminase. The structure of glucagon has been reviewed by Blundell and Wood (1982).

A number of glutamine-containing polypeptide hormones do not have

NH_2H-S-Q-G-T-F-T-S-D-Y-S-K-Y-L-D-S-R-R-A-Q-D-F-V-Q-W-L-M-N-T-COOH

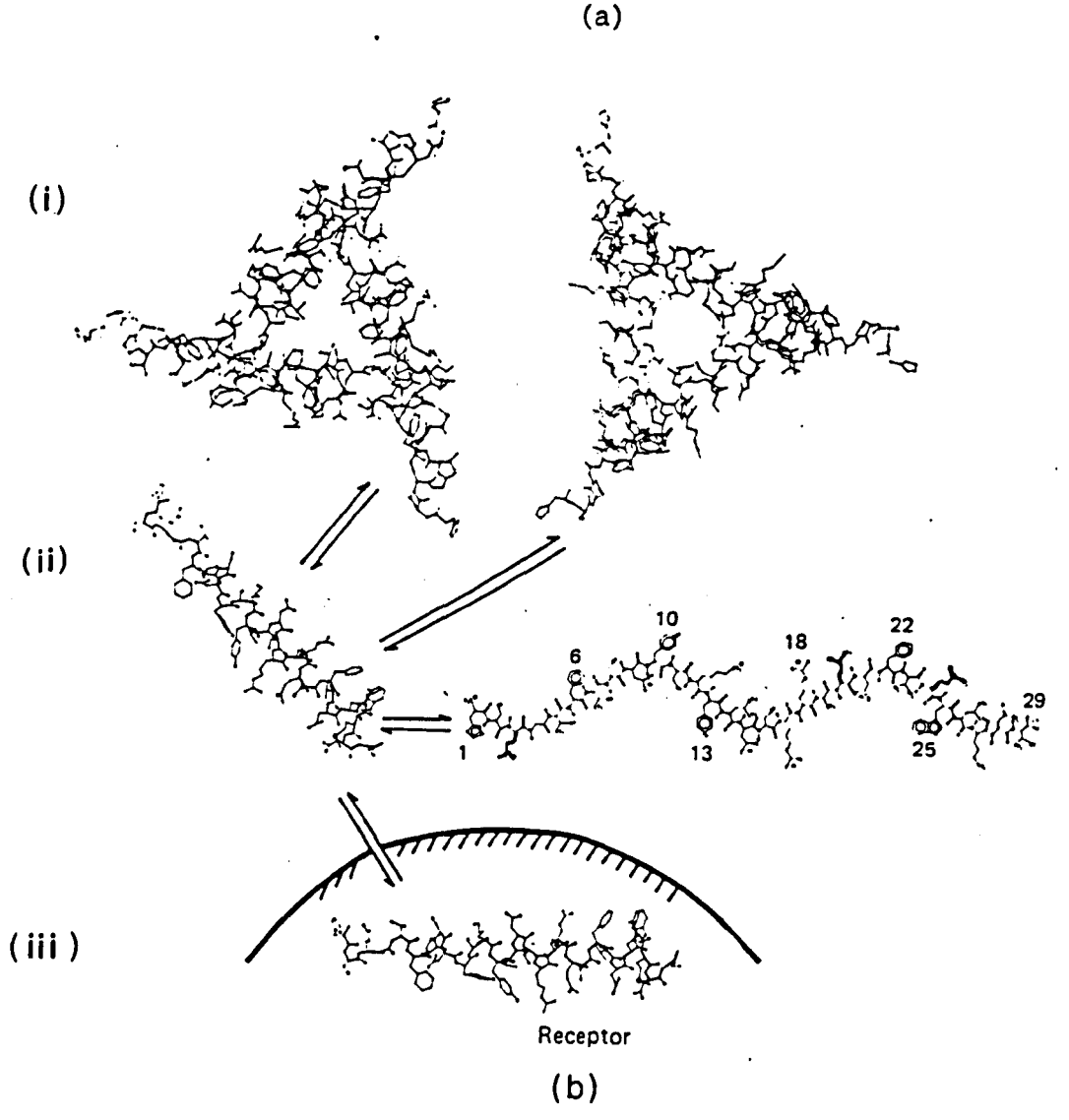


Fig 1.7. THE STRUCTURE OF BOVINE GLUCAGON

(a) The primary amino acid sequence

(b) (i) The differently folded trimeric structures of glucagon.
(ii) The monomeric form in helical and an extended conformation which are thought to exist in dilute aqueous solution.

The trimeric forms are thought to dissociate to give an eqilibrium population of flexible monomers, but at the receptor an α-helical conformer (iii) is probably stabilized. Diagrams (b) is from Blundell (1979).

reported that the cyclic peptides oxytocin, and vasopressin, show no substrate characteristics. Their primary sequences are shown in Fig 1.8.

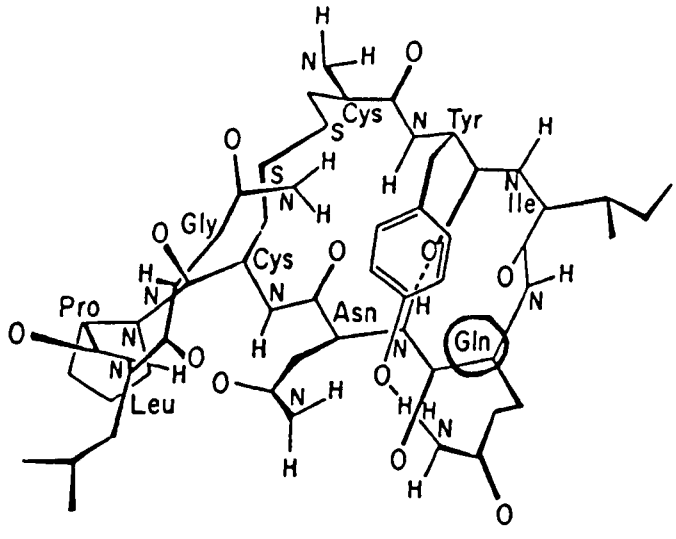
Fig 1.8. THE AMINO ACID SEQUENCE OF a) VASOPRESSIN AND b) OXYTOCIN [C⌐¬C = cystine].

Although these cyclic peptides are known to be highly flexible in aqueous solution (Urry et al., 1981), they have reduced conformational properties relative to their linear counterparts. Their non-linear topography may exclude their entry at the active site of transglutaminase if preferred substrates bind in extended or α-helical forms. The folded structure of these peptides based on nmr studies (Urry et al., 1981) is shown in Fig 1.9.

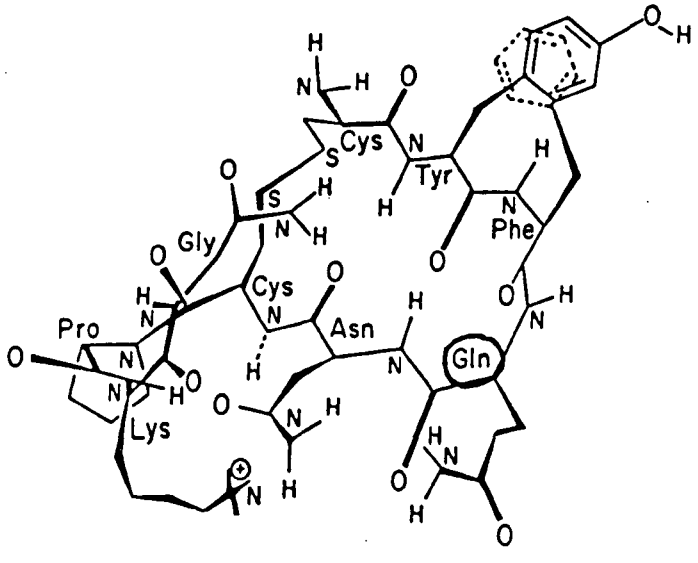
1.7.7. DENATURED POLYPEPTIDES AND PROTEINS

1.7.7.1.INTRODUCTION

The following sections (1.7.7.2. to 1.7.7.3.) consider the evidence concerning the differential reactivity of glutaminyl residues within denatured polypeptide sequences which contain relatively large numbers



(a) THE STRUCTURE OF VASOPRESSIN



(b) THE STRUCTURE OF OXYTOCIN

Fig 1.9. THE NMR STRUCTURES OF TWO UNREACTIVE NEUROPHYSEAL HORMONES

The similar conformations of vasopressin (a) and oxytocin (b) are shown in their proposed biologically active conformations above. Viewed from the top there is a hydrophil ic cluster of amino acid side chains in each molecule. In vasopressin this comprises the hydroxyl of Tyr_2 and the carboxamide groups of Gln_4, Asn_5, and $Glycinamide_9$ groups. In oxytocin the cluster comprises Gln_4, Asn_5, $glycinamide_9$ and the basic moiety of Lys_8. The sequences of these peptides are shown in Fig 1.8 and in solution they are thought to be highly flexible molecules. The diagrams are taken from Urry et al., 1981.

of highly exposed glutaminyl residues with diverse flanking sequences surrounding them. In section 1.7.2. it was stated that most glutaminyl residues within proteins are potentially reactive given sufficiently forcing conditions eg. long incubation times in the presence of high concentrations of transglutaminase. However by reducing incubation times and using reduced quantities of enzyme the most reactive glutaminyl residues may be identified. The amino acid sequences surrounding reactive sites may be compared with substrate sites within synthetic substrates (see section 1.7.5.) and polypeptide hormones (see section 1.7.6.) in search of consensus features. It is also possible to examine the sequence motifs associated with unreactive sites for features which may discourage reactivity of these sites.

1.7.7.2. DENATURED HAEMOGLOBIN AND INSULIN

The sites of modification by transglutaminase within heat denatured haemoglobin (Pincus and Waelsch 1968) and oxidized insulin (Toda and Folk, 1965) are aligned against the reactive glutaminyl residue in Table 1.2.. Samples were incubated for short incubation times (1 to 4 hr) and this approach revealed a number of preferred sites of labelling.

A consideration of the primary structure surrounding reactive sites indicates a diversity of sequence comparable with that noted in polypeptide hormone and synthetic peptide substrates (see sections 1.7.5. and 1.7.6.).

The negative charge (pKa = 1.8) imparted by oxidation of free cysteine side chains to the sulphonic acid derivative at positions : -4, +1, +2, +3,

+5 did not prevent reaction of neighbouring glutamine residues within insulin (Toda and Folk, 1965).

Table 1.2. SUBSTRATE SITES FOR TRANSGLUTAMINASE WITHIN DENATURED INSULIN AND HAEMOGLOBIN

PROTEIN	RESIDUE No	AMINO ACID SEQUENCE	RESIDUE No
Hb A	48	L-S-H-G-S-A-Q-V-K-G-H-G-K-K	61
Hb A	121	E-F-T-P-P-V-Q-A-A-Y-Q-K-V-V	134
INS A	9	-V-C_X-S-L-Y-Q-L-E-N-Y-C_X-	22
INS A	1	NH_2-G-I-V-E-Q-C_X-C_X-A-S-V-C_X	11
INS B	1	NH_2-F-N-N-Q-H-L-C-G-S-H-L	11

Key: Hb= Human haemoglobin A= -chain, B= B-chain. INS= Human insulin [periodic acid treated ($_X$=sulphonic acid derivative of cysteine residue)]. Δ = positively charged residues and O = negatively charged residues.

The positive charge imparted by the lysyl residues on the C-terminal side of Gln_{54} in haemoglobin may reduce the reactivity of this and other sites, since succinylation of haemoglobin greatly improved the substrate properties of denatured haemoglobin, Pincus and Waelsch, (1968). The unreactive sites within non succinylated haemoglobin are aligned in Fig 1.10.

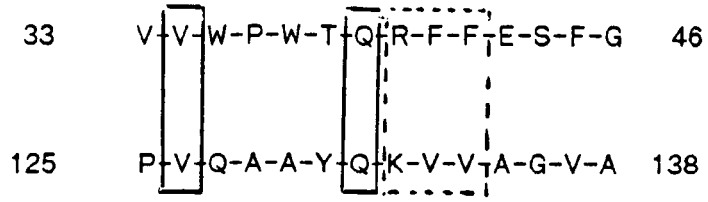


Fig 1.10. UNLABELLED GLUTAMINYL RESIDUES OF HUMAN HAEMOGLOBIN (β-CHAIN) Residues at identical positions are boxed with a solid line , while other similar motifs are boxed ith hatched lines

The presence of the positively charged arginine and lysine at position +1 may reduce substrate properties of these sites. However the presence of a number of bulky side chains within these sequence could equally cause this by steric blocking eg. Trp_{37}, Phe_{41}, Phe_{42}, and Tyr_{130}. Although these proteins are considered denatured following treatment with heat and/or chemical modification, some local secondary structure may be retained in regions adjacent to potential glutaminyl substrate sites. The secondary structure associated with some of these sites has been predicted using computer based methods, see section 8.2.. It may also be significant that unreactive Gln_{39} has a -Q-R-F-F- motif which resembles the -Q-Q-F-F- motif associated with the reactive Gln_5 in substance P (see section 1.7.6.2.) altough the reactive sequence lacks positive charge.

1.7.7.3. SUCCINYLATED BOVINE β-CASEIN

a) INTRODUCTION

Bovine β-casein has a largely disorganized structure at neutral pH (Herskovitz, 1966). The amount of discretely folded structure is thought to be reduced further at high pH when in its succinylated form (Yan

to be reduced further at high pH when in its succinylated form (Yan and Wold, 1984). The structure therefore is assumed to approximate a random coil in this form. The majority of the glutaminyl residues are expected to be fully exposed. While this assumption may exaggerate the true solution structure of β-casein, the molecule is nevertheless a suitable model system for the investigation of determinants of transglutaminase specificity which reside within the linear sequence of polypeptide and protein substrates. The rationale for this is that;

1) Large stretches of the molecule may approach a random coil conformation at any one time. Thus it must be features of the linear sequence of succinylated (and hence denatured) β-casein, and not gross conformational constraints which determine the reactivity of glutaminyl residues†.

2) β-casein contains a total of twenty two glutaminyl residues, which are surrounded with a diverse distribution of charged, uncharged, hydrophobic, hydrophilic etc. amino acid residues. Hence by comparing reactive and non-reactive sequences within this substrate a large amount of information concerning transglutaminase specificity may be deduced from a single protein substrate.

† In the absence of stable secondary structure, steric blocking of glutaminyl residues from the active site of transglutaminase may be facilitated by the presence of large amino acid side chains (eg. tryptophan, isoleucine etc.) if they are sufficiently close to a glutaminyl residue even if relatively free rotation is possible at the peptide bond linking the two residues.

b) THE PRIMARY SEQUENCES ASSOCIATED WITH SUBSTRATE SITES IN SUCCINYLATED β-CASEIN

The modification sites within the succinylated protein were determined by direct sequencing (Yan and Wold, 1984), using dansylcadaverine and N-(Glc-Glc-glucitol-1)-cadaverine [maltotriose reductively amidated with cadaverine as acyl acceptors. The labelled sites are shown in Table 1.3..

Table 1.3. THE SUBSTRATE SITES FOR TRANSGLUTAMINASE WITHIN SUCCINYLATED BOVINE β-CASEIN

AMINO ACID SEQUENCE *		
RESIDUE NUMBER		RESIDUE NUMBER
47	D-K-I-H-P-F-A-Q-T-Q-S-L-V-Y-P	61
49	I-H-P-F-A-Q-T-Q-S-L-V-Y-P-F-P	63
72	P-Q-I-P-P-L-T-Q-T-P-V-V-V-P-P	86
161	Q-S-V-L-S-L-S-Q-S-K-V-L-P-V-P	175
168	S-K-V-L-P-V-P-Q-K-A-V-P-Y-P-Q	182
175	Q-K-A-V-P-Y-P-Q-R-D-M-P-I-Q-A	189
187	I-Q-A-F-L-L-Y-Q-Q-P-V-L-G-P-V	200

* Amino acid residues which are present at identical positions relative to the modified glutaminyl residues within these sequences have been boxed.

All the labelled sites were modified to an extent > 80% except for Gln_{54} and Gln_{182}, which were modified to 30 % and 60 % respectively. The labelled sites tend to be in relatively hydrophobic sequences and lacking in charged residues relative to many of the unlabelled residues (see Table 1.4.). It may be significant that positively charged residues are generally not present in the region C-terminal to reactive glutaminyl residues. In those cases where they are present the charge on the lysine residue had

been reversed by succinylation, or the glutaminyl residue was only partially modified indicating a reduced reactivity.

c) THE SECONDARY STRUCTURAL POTENTIAL ASSOCIATED WITH GLUTAMINYL SITES

A secondary structural prediction of the structure of bovine ß-casein is available (Holt and Sawyer, 1988). None of the labelled glutaminyl residues lie within areas predicted to be high in α-helix or ß-sheet potential. Two lie within region predicted to be on or near ß-turns, while the rest lie in regions of no predicted structure. Since the glycosyl groups of the amine substrate [(Glc-Glc-glucitol)cadaverine] were rather bulky and the glutaminyl residues Gln_{52} and Gln_{54} close to each other, the reduced labelling of Gln_{54} was probably due in part to steric blocking by modified Gln_{52}. The lack of substrate properties of many of the sites shown may be due to their lack of accessibility to the enzyme.

Yan and Wold, (1984) have noted the high content of proline and hydroxylated amino acid residues which contribute towards the high ß-turn potential associated with the labelled sites. Space filling models of Gln-Ser, Ser-Gln, Gln-Tyr, and Tyr-Gln, indicate that the hydroxyl group of the serine or tyrosine can be brought close enough to interact with the amide group of the glutaminyl side chain so long as the dipeptide is not involved in α-helix or ß-sheet structures. Such interactions could increase the reactivity of glutaminyl residues (see section 10.5.(5)..

d) UNREACTIVE GLUTAMINYL SITES IN β-CASEIN

The unreactive glutaminyl sites are aligned in Table 1.4.. The sequences tend to either be characterized by having a high charge density about the glutaminyl residue, and/or a high frequency of occurrence of proline residues and bulky amino acid side chains in positions neighbouring the unreactive glutaminyl sites eg. tryptophan, phenylalanine and leucine. It may be significant that there is no published example of a reactive glutaminyl site having an adjacent proline residue on the C-terminal side, or a leucine residue on its N-terminal side.

Table 1.4. ALIGNMENT OF THE UNREACTIVE GLUTAMINYL SITES WITHIN SUCCINYLATED BOVINE β-CASEIN

RESIDUE No	AMINO ACID SEQUENCE *	RESIDUE NO
27	[K-K]-I-(E)-[K]-F-[Q]-S-(E)(E)-Q-Q-Q	39
32	[K]-F-Q-S-(E)(E)-[Q]-Q-Q-T-(E)(D)(E)	44
33	F-Q-S-(E)(E)-Q-[Q]-Q-T-(E)(D)(E)-L	45
34	Q-S-(E)(E)-Q-Q-[Q]-T-(E)(D)(E)-L-Q	46
41	Q-T-(E)(D)(E)-L-Q-(D)[K]-I-H-P-F	53
66	I-P-N-S-L-P-[Q]-N-I-P-P-L-T	78
83	V-V-P-P-F-L-[Q]-P-(E)-V-M-G-V	95
110	P-F-P-[K]-Y-P-[Q]-P-F-T-T-(E)-S	122
117	Q-P-F-T-(E)-S-[Q]-S-L-T-L-T-(D)	129
135	L-P-P-L-L-L-[Q]-S-W-M-H-Q-P	147
140	L-Q-S-W-M-H-[Q]-P-H-Q-P-L-P	152
143	W-M-H-Q-P-H-[Q]-P-L-P-P-T-V	155
154	T-V-M-F-P-P-[Q]-S-V-L-S-L-S	166
182	Q-[R]-(D)-M-P-I-[Q]-A-F-L-L-Y-Q	194
189	A-F-L-L-Y-Q-[Q]-P-V-L-Q-P-V	201

* Positively charged residues are boxed and negati vely charged residues are in circles.

Many of the unreactive glutaminyl residue containing sequences are predicted to have high secondary structure forming potential (see section 9.2.3.). This may restrict flexibility at these sites relative to more reactive sites, and hence reduce access of these unreactive glutaminyl residues at the active site of transglutaminase.

The lack of reactivity of glutaminyl residue: Gln_{123} is the most difficult result to explain since the surrounding sequence is similar in sequence to

reactive sites in β-casein (eg Gln_{56} and Gln_{167}) and a number of other protein substrates eg β-endorphin (see section 1.7.6.2.). However it may be significant that although a total of seven substrate sites were determined within β-casein, the maximum observed stoichiometry of modification of the protein using dansylcadaverine as substrate was 8.0 mol amine/mol protein monomer. Hence it is possible that Gln_{123} does represent a substrate site, but was not identified during the characterization of the many modified products (Yan and Wold, 1984 a).

1.7.8. SUBSTRATE SITES WITHIN GLOBULAR PROTEINS

1.7.8.1. INTRODUCTION

Globular proteins are generally poor substrates for transglutaminase compared to non-globular or denatured proteins (Toda and Folk, 1969). If a globular protein does have substrate properties for transglutaminase, it is unusual for more than two glutaminyl residues within the structure to be reactive (Brenner and Wold, 1978). The reduced substrate properties of globular proteins compared to their more flexible denatured counterparts suggests that certain types of secondary structure may make otherwise reactive glutaminyl residues almost totally unreactive when the protein is in its native state.

Globular proteins, which have substrate properties for transglutaminase are useful in the construction of theories concerning the specificity of transglutaminase. This is particularly so when physicochemical models of them are available eg. based on X-ray crystallographic and n.m.r. studies. In these cases it is possible to assess the local conformation associated with reactive and unreactive glutaminyl residues. In those cases where

local conformation is not expected to cause reduced substrate properties within a glutaminyl site, it is assumed that features associated with the linear amino acid sequence are important in this regard.

1.7.8.2. FEATURES ASSOCIATED WITH PRIMARY SEQUENCES

The amino acid sequences surrounding the known substrate sites within globular substrates are aligned in Table 1.5.

TABLE 1.5. SUBSTRATE SITES FOR TRANSGLUTAMINASE WITHIN GLOBULAR PROTEINS

PROTEIN	RESIDUE NO	AMINO ACID SEQUENCE	RESIDUE NO
ßBpcrystallin	2	S-N-H-(E)-T-Q-A-G-[K]-P-Q-P	13
ßB3crystallin	3	NH_2-[A]-(E)-Q-H-S-T-P-Q-Q	9
ßA3crystallin	1	NH_2-[E]-T-Q-T-V-Q-Q-(E)-L-(E)-S	10
lipocortin I	12	F-I-(E)-N-(E)-(E)-Q-(E)-Y-N-Q-T	23
phospholipase A2	1	NH_2-[A]-L-W-Q-F-[R]-S-N-I	9
actin	33	I-V-G-[R]-P-[R]-H-Q-G-V-M-V-G	45
SV-1V	3	T-[K]-(E)-[K]-Y-S-Q-S-(E)-(E)-V-V-S	15
SV-1V	79	[K]-[K]-[K]-S-[R]-F-A-Q-(D)-V-L-(N)-COOH	90
collagen III	8	G-C-S-H-L-G-Q-S-Y-A-(D)-[R]	19

positively charged residues are enclosed in triangles and negatively charged residues are enclosed in circles. N- and C- terminal residues can be in both catagories simultaniously REFERENCES ; crystallins: Berbers et al., (1984); lipocortin I; Pepinsky et al., (1989); mellatin Miele et al., (1990); actin; Takashi, (1988); SV-IV : Porta et al., (1991); collagen: Bowness et al., (1987)

The sequences about reactive glutaminyl residues in globular proteins appear to be as diverse as those associated with conformationally less

restricted substrates (see sections 1.7.3. to 1.7.7.). Several of the substrates are rich in negative charge in the region surrounding the modification site. In others the surrounding sequences are reminiscent of those present in other categories of transglutaminase substrate eg. the cluster of positive charge N-terminal to Gln_{86} in rSBP is similar to that seen in melittin (see section 1.7.6.2.). One difference between native globular substrates and other more flexible substrates is the presence of positive charge in the region C-terminal to the labelled glutamine in certain cases eg. Arg_6 of phospholipase A2.

1.7.8.3. CONFORMATIONAL FEATURES ASSOCIATED WITH SUBSTRATE SITES

The tertiary structure of some of some substrate proteins have been determined and the secondary structures associated with the modification sites are diverse.

1) Gln_{41} of actin is located on the N-terminal side of a fourteen residue surface loop which is bounded by β-sheet on either side (Kabsch et al., 1990), [see Fig 1.11].

2) Gln_4 of porcine phospholipase A2 (see Fig 1.5) is located within an N-terminal α-helix (Dijkstra et al., 1983), [see Fig 1.12].

3) Gln_7 of β-Bp-crystallin is located in an extended region lacking electron density in the model of Bax et al., (1990), [see Fig 1.13].

In other cases the secondary structure has been assigned on the basis of either: 1) known similarity of folding of a protein to one of known X-ray structure [eg. the small basic protein of rat testis (rSBP) is known to resemble the globular protein, uteroglobulin, (Morize et al., 1987) both in sequence (Menne et al., 1982) [see Appendix III] and immunologically, (Metafora et al., 1987)], 2) the results of algorithms designed to predict secondary structures. From this the following structures are inferred:

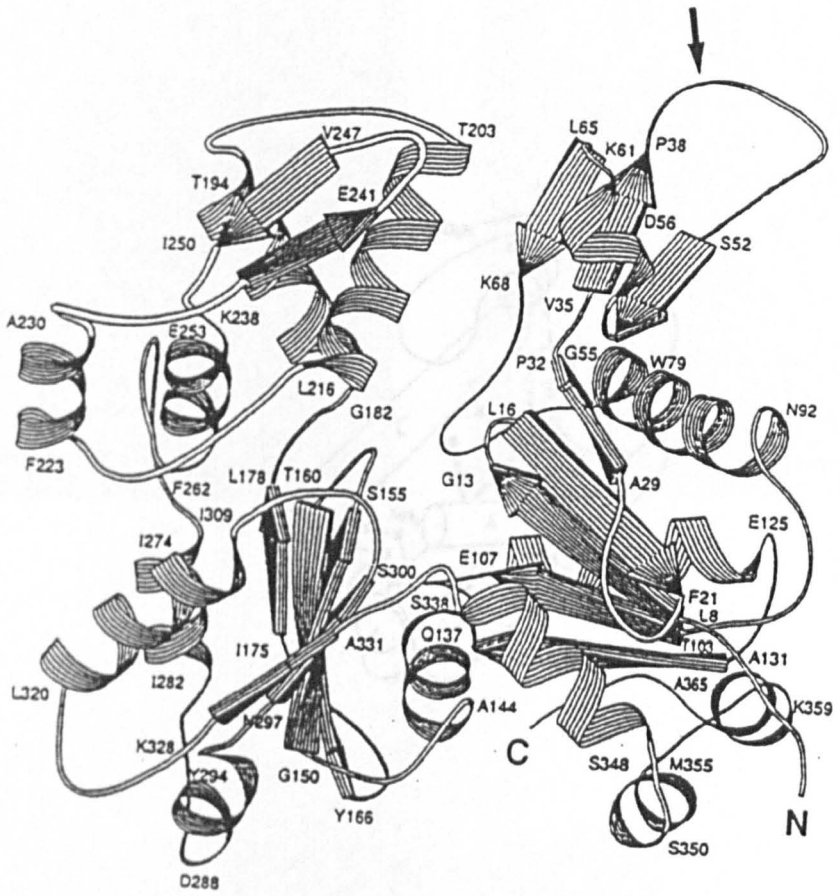


Fig 1.11. SCHEMATIC DRAWING OF THE X-RAY STRUCTURE OF ACTIN FROM RABBIT MUSCLE.

Regions of ß-sheet are arrow-like bars and α-helix are tubes. The numbering represents the first and last residues of the amino acid sequence corresponding to each region. The protein was co-crystallized with DNAase (Kabsch et al., 1990) and hence this interaction may affect the folding of the actin monomer. The position of the surface loop within which Gln_{41} is located is indicated by an arrow, (⟵).

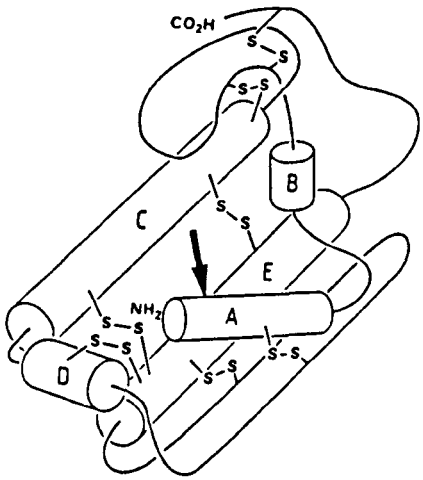
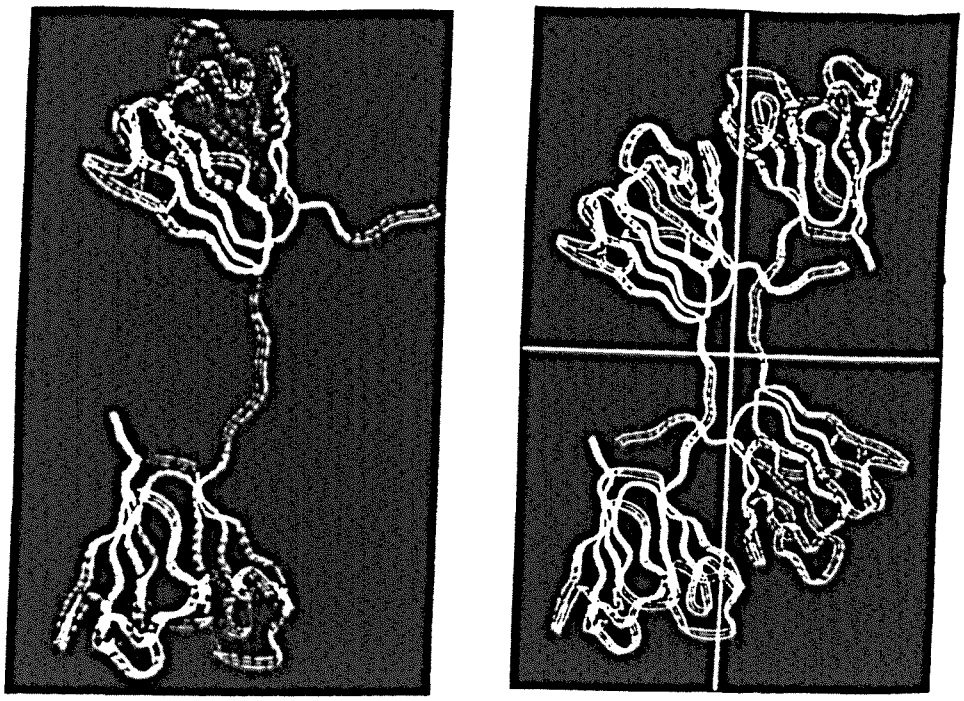


Fig 1.12. SCHEMATIC DRAWING OF THE X-RAY STRUCTURE OF PORCINE PHOSPHOLIPASE A2.

Regions of α-helix are depicted as tubes. The structure shown above is that of bovine phospholipase, but this is known to have a very similar overall tertiary structure to the porcine enzyme (Dufton et al., (1983). The helix nomenclature is that of Dijkstra et al., (1978). There is only one glutamine within the primary sequence, and this reactive glutaminyl (Gln_4) is located in helix A at the amino terminus and this is indicated with an arrow. See Fig 1.15 for a diagram of the stereochemistry of the amino acid side chains in the region surrounding Gln_4.



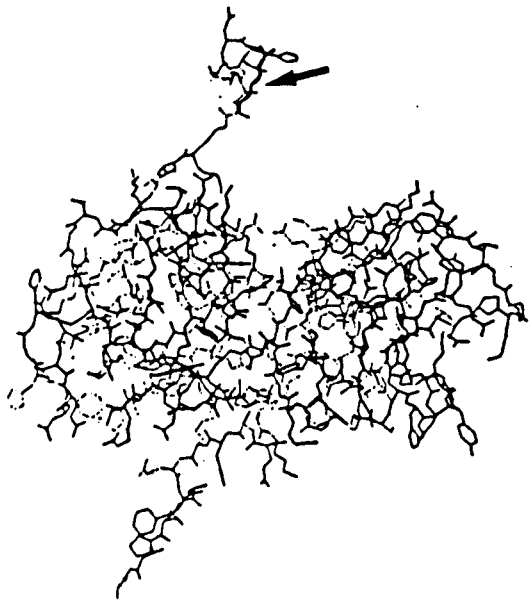


Fig 1.13. THE CRYSTAL STRUCTURE OF BOVINE β-Bp CRYSTALLIN.

The monomeric model of the bovine β-Bp crystallin (Bax et al., 1990) is shown schematically in extended form in diagram (a). This shows clearly the "connecting peptide", which lies between domains and the N- and C- termini. Diagram (b) shows the dimeric arrangement of the protein. The tertiary structure is shown in greater atomic detail in diagram (c) where the reactive Gln_7 which is shown projecting upwards in bold type and is marked with an arrow.

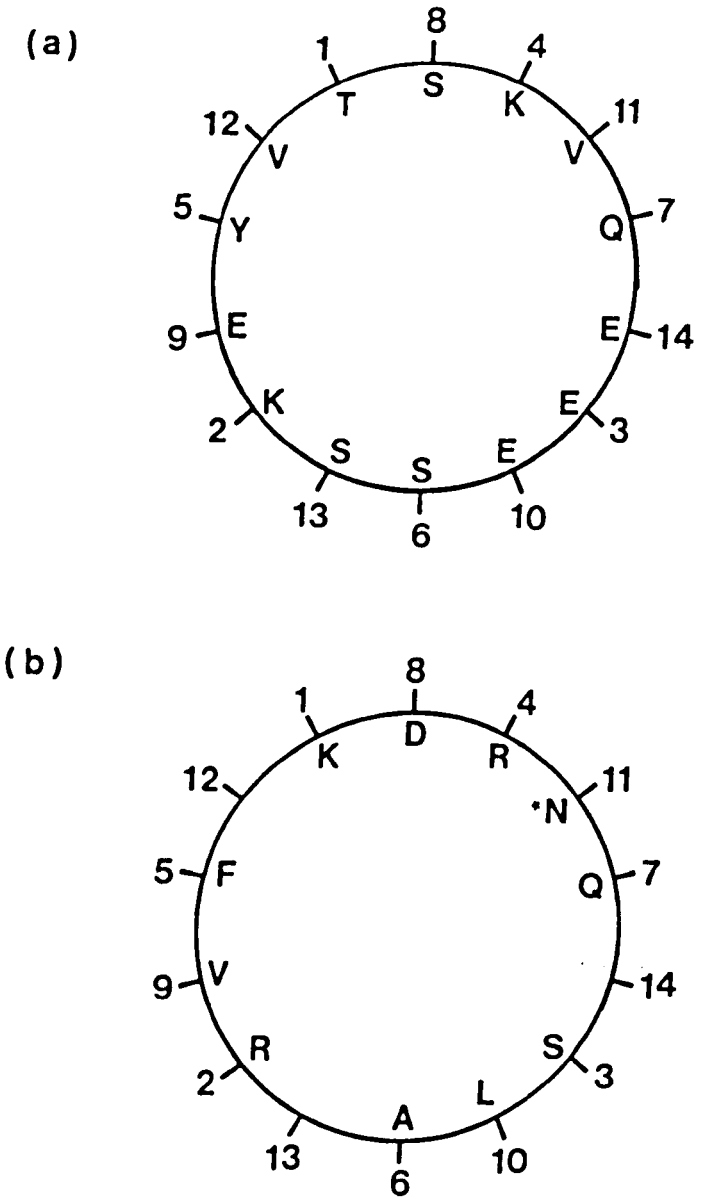


Fig 1.14. HELICAL WHEEL REPRESENTATIONS OF THE SEQUENCES SURROUNDING THE REACTIVE GLUTAMINYL RESIDUES OF RAT SV-IV PROTEIN.

(a) The region surrounding glutaminyl residue Gln_9.

(b) The region surrounding glutaminyl residue Gln_{86}.

In both cases negatively charged residues are predicted to be in close spatial proximity to the reactive glutaminyl residue. This phenomenon is also observed in lipocortin I, β-Bp-crystallin, and α-A2 crystallin. It may also be significant that other glutaminyl residues within substrate sequences are often found clustered about the reactive residue when they are modelled in this way. In positions where no amino acid residue is indicated, the polypeptide chain has terminated prior to these positions.

[Symbols; N* = C-terminal Asn residue]

1) Glutaminyl residue Gln_{18} of human placental lipocortin I is predicted to be located within a N-terminal α-helix [see section 9.2.3.]

2) Glutaminyl residues Gln_{9}, and Gln_{86} of SV-IV protein from rat are predicted to be located within N-, and C- terminal α-helical regions of the molecule. These are shown as helical wheel representations in Fig 1.14.

The local secondary structure associated with Gln_{14} of bovine collagen is not known, but is conformationally restricted by disulphide bonds, (Brandt et al., 1984). The conformational restriction appears to favour reactivity at this site, since reduction of these disulphide bond has been reported to reduce the reactivity of Gln_{14}. This is discussed further in section 1.7.9.4..

1.7.8.4. FEATURES ASSOCIATED WITH UNREACTIVE GLUTAMINYL RESIDUES

Many globular protein substrates of transglutaminase appear to contain exposed glutaminyl residues which are either unreactive or have poor substrate properties towards transglutaminase. Some of the sequences associated with these sites are shown in Table 1.6. In all of these cases

the unreactive glutaminyl residues are located within the same region of secondary structure as the reactive site(s) shown in Table 1.5.

Table 1.6. UNREACTIVE GLUTAMINYL SITES WITHIN GLOBULAR PROTEIN SUBSTRATES OF TRANSGLUTAMINASE

PROTEIN	RESIDUE NUMBER	AMINO ACID SEQUENCE
B-Bp crystallin	(Q12)	T-Q-A-G-K-P-Q-P-L-N-P-K-I-I
B-B3 crystallin	(Q3)	NH_2-E-T-Q-T-V-Q-Q-E-L-E
lipocortin I	(Q22)	E-E-Q-E-Y-V-Q-T-V-K-S-S-K-G
lipocortin I	(Q10)	V-S-E-F-L-K-Q-A-W-F-I-E-N-E
actin	(Q51)	V-M-V-G-M-G-Q-K-D-S-Y-V-G-D
collagen	(Q1)	PCA-Q-E-A-V-D-G-G-C

Abbreviation; [PCA=pyrrolidone carboxylic acid]

Positively charged residues are in triangles and negatively charged residues are in circles

It is not clear whether the poor substrate properties of the unreactive glutaminyl residues shown in Table 1.6. are due to either;

1) a lack of conformational flexibility about these sites

2) steric factors within the primary sequences, or,

3) chemical features such as the juxtaposition of unfavourable charged residues about these sites.

A theoretical basis for the observed lack of reactivity of these residues is discussed further in section 1.7.9.

1.7.9. THE SPECIFICITY OF TRANSGLUTAMINASE : A DISSECTION OF CONFORMATION FROM SEQUENCE DETERMINANTS OF SPECIFICITY

1.7.9.1. INTRODUCTION

By comparing the amino acid sequences associated with substrate sites in globular proteins (see section 1.7.8.) with those of non-globular substrates (see sections 1.7.3. to 1.7.7.) the effect of conformational restriction on the distribution of amino acid residues within substrate sites may be qualitatively evaluated. Within a peptide which is highly flexible in solution, there should be an approximation of free rotation about many of the constituent peptide bonds relative to each other. Hence residues which both increase or decrease the reactivity of a glutaminyl residue by their proximity will have ample opportunity to affect a potential substrate site.

Inhibitory residues should be conspicuous by their absence in effective substrates for transglutaminase. However in more conformationally restricted substrates, residues which are inhibitory to the activity of transglutaminase will not be expected to approach closely to reactive glutaminyl residues, while residues which improve substrate properties will be expected to be present in higher proportions than in the less discriminating flexible substrates. By using this approach the criteria which appear to determine the specificity of transglutaminase are grouped into a hierarchy of levels in sections 1.7.9.2. to 1.7.9.4..

1.7.9.2. FLEXIBILITY ABOUT GLUTAMINYL SITES

In the absence of unfavourable charge or steric effects, flexibility about substrate sites appears to be a feature which is usually associated with

non-physiological substrate sites within proteins. In the globular substrates tested up to December 1990, the modification sites of transglutaminase are located on the N-termini or large flexible loops. There is evidence however that flexibility about substrate sites is not an essential requirement for substrate properties. For example :

1) The X-ray structure of phospholipase A2 indicates (in the crystallized form) that Gln_4 is located within an α-helix which extends from Ala_1 to Glu_{11} (de Haas et al., 1987). This region has been predicted to have strong helix forming potential (Dufton et al., 1983) although in solution the same region has been observed from n.m.r. studies to be somewhat less ordered (Dekker et al., 1991). This site is not predicted to bind at the active site of transglutaminase in an extended form (Fig 1.5), since by analogy with other highly flexible substrates (see sections 1.7.2. to 1.7.7.) the positive charge imparted by Lys_6 would be expected to reduce the substrate properties of Gln_4 unless in a folded conformation.

2) The substrate site within collagen appears to benefit from some conformational rigidity associated with its disulphide bonded state. The cleavage of disulphide bridges in the collagen propeptide by DTT was observed to abolish the reactivity of Gln_{12} towards transglutaminase, (Bowness et al., (1987). DTT treatment is likely to increase conformational flexibility about Gln_{12}. This illustrates the probable importance of chemical factors within the linear sequence of protein/peptide substrates.

3) The majority of exposed glutaminyl residues in globular proteins are unreactive towards transglutaminase. This suggests that the extended substrate binding site of transglutaminase (Folk, 1980) requires not just exposure of the glutaminyl residue but also a sufficiently long region

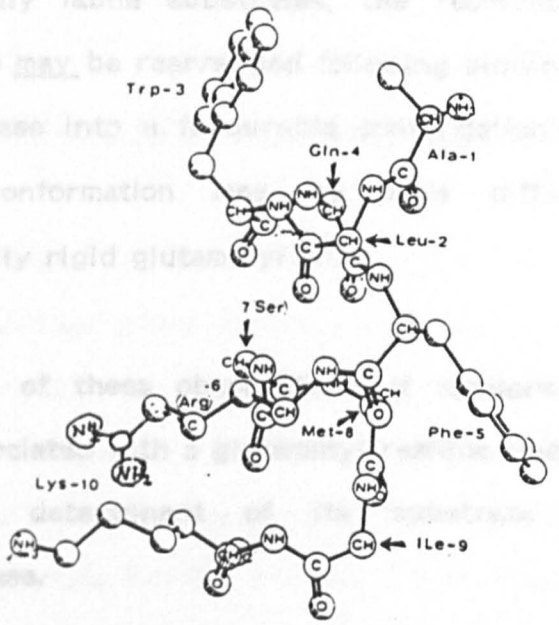


Fig 1.15. THE STRUCTURE OF THE N-TERMINAL HELIX OF PORCINE PHOSPHOLIPASE A2

The computer drawing is from De Haas et al., (1987). For the sake of clarity only the side chains of Ala_1, Trp_3, Asn_6, Lys_{10} are shown. The side chain of Gln_4 which is reactive towards transglutaminase will project into the page and should be close in space to Ala_1, and Ser_7. The large hydrophobic side chains of Trp_3 and Phe_5 are shown to project away fom the page and hence are not in a position to sterically block access of transglutaminase to the reactive site.

Note that both of the positively charged residues in the primary sequence surrounding Gln_4 ie. Arg_6 and Lys_{10} are facing out of the page and are thus spatially remote from the site of modification

surrounding the glutaminyl residue to allow multiple (possibly cooperative) interactions, before reaction can occur. It is possible that transglutaminase may not have such a requirement for flexibility if a glutaminyl residue (and possibly the side-chains of surrounding residues) within a polypeptide sequence are folded into a conformation which resembles the transition state of the enzyme-substrate complex. In conformationally labile substrates, the residues within an extended substrate site may be rearranged following binding to the active site of transglutaminase into a favourable conformation for reaction. Such a favourable conformation may be more difficult to achieve at conformationally rigid glutaminyl sites.

On the basis of these observations it appears that the secondary structure associated with a glutaminyl residue must be considered to be a LEVEL 1 determinant of its substrate properties towards transglutaminase.

1.7.9.3. SHORT RANGE STERIC FACTORS

Although some regions within globular proteins appear to be both flexible and lacking in charged residues which might reduce substrate properties they remain unreactive. The best explanation for this is that the presence of bulky side-chains may sterically block access of transglutaminase to these sites. Many of the features associated with linear sequences which surround these unreactive sites in globular proteins are also common to the unreactive sites in β-casein [see section 1.7.7.3.(c),]. For example: bulky side chains such as tryptophan, leucine, methionine and proline residues neighbour glutaminyl residues. There are many other substrates

which illustrate this point eg, note that Gln_{12} in β-Bp crystallin has proline residues at positions -1 and +1 (see Table 1.6.).

The ability of neighbouring residues to restrict the local conformation of a glutaminyl residue may be considered a LEVEL 2 † (see footnote of section 1.7.9.) determinant of transglutaminase specificity.

1.7.9.4. THE DISTRIBUTION OF CHARGED RESIDUES

Positively charged residues are present in the region C-terminal to glutaminyl residues which have been shown to be poorly reactive eg. human haemoglobin, bovine β-casein. Charged residues are usually presented on the surface of proteins and therefore their low frequency within substrate sites could explain why many globular proteins have poor substrate properties for the enzyme. Thus although in some cases the observed lack of reactivity is probably due to unfavourable local conformation, a possible explanation of the lack of reactivity other sites within less conformationally constrained proteins is because of the unfavourable distribution of positively charged residues. If such features do underlie the basis of transglutaminase specificity, then the distribution of negatively charged residues could also be important in the determination of the substrate properties of a given glutaminyl residue (see chapter 10.).

The presence or absence of charged residues at certain unfavourable positions within the primary sequence of transglutaminase substrates may

therefore be considered to be a LEVEL 3 † (see footnote of section 1.7.9.) determinant of transglutaminase specificity.

1.7.9.5. HYDROXYLATED AMINO ACIDS

The presence of hydroxylated amino acids in the sequences bordering substrate sites is a common motif associated with glutaminyl reactivity. Since this feature is not seen in every case, but appears to improve substrate characteristics this may be considered a LEVEL 4 determinant of transglutaminase specificity. Other features could equally be assigned to this level include:

1) a high frequency of glycine residues within substrates,

2) the presence of a Val-Leu (or related sequence) at positions +3 and +4.

Such motifs may be as effective by increasing conformational flexibility (see section 1.7.9.2.) as by any other chemical effects. However hydroxylated amino acids are strong hydrogen bond formers, and this could also account for their presence within substrate sites.

† ; It should be noted that in some cases charged residues (level 3 determinants) may act as level 2 determinants of transglutaminase specificity through steric blocking of glutaminyl residues, independently of their associated charge effects on substrate properties.

1.8. THE SPECIFICITY OF FACTOR XIIIa TOWARDS PROTEIN SUBSTRATES

1.8.1. INTRODUCTION

A number of substrate sites of factor XIIIa have been determined and are shown aligned against each other in Table 1.7. (see over page). These sites are compared and discussed in relation to what is known concerning the known modification sites of the guinea pig liver enzyme (see section 1.7).

1.8.2. PROTEIN SUBSTRATES OF FACTOR XIIIa

All of the protein substrates of factor XIIIa listed in Table 1.7. were modified in their native forms except bovine rhodopsin which was succinylated, prior to its modification by transglutaminase.

A -Q-Q- motif is apparent in a number of the substrate sites within Table 1.7., and is particularly associated with those proteins which are thought to be substrates for factor XIIIa *in vivo*. This motif may confer favourable properties to substrate sites within proteins which are not thought to be physiological substrates of Factor XIIIa. It may be significant therefore that the -Q-Q-Q- motif within bovine rhodopsin has favourable substrate properties for the enzyme. Apart from this motif there is (like the liver enzyme) considerable diversity about modification sites. Thus it is not surprising that some of these sites may also be reactive towards the guinea pig liver enzyme eg. fibrin and fibronectin (see section 1.8.4.).

In all cases the substrate sites are expected to be within flexible regions of polypeptide chain. The secondary structure associated with the substrate site of factor XIIIa within α_2-macroglobulin has been

by n.m.r., (Gettins et al., 1988) and predicted to be a flexible region of parallel ß-sheet within the so called "bait" region. The structure of α_2-macroglobulin has been reviewed by Van Leuven (1982). The membrane orientation of rhodopsin has been investigated by a number of different methods [reviewed by Ovchinnikov (1987)] and is shown in Fig 1.16. Little is known about the folded structure of the other substrates of factor XIIIa.

Table 1.7. THE REACTIVE SITES WITHIN FACTOR XIIIa SUBSTRATES

PROTEIN	RESIDUE	AMINO ACID SEQUENCE	RESIDUE
FIBRIN	322	G-T-G-S-T-G-H-**Q**-N-P-G-S-P-R	334
FIBRIN	359	S-V-S-G-S-T-G-**Q**-W-H-S-E-S-G	372
FIBRIN	390	R-L-T-I-G-E-G-**Q**-Q-H-H-L-G-G	403
VITRO	87	-S-K-G-N-P-E-**Q**-T-P-L-V-K-P	99
α_2PI	1	A-Q-E-**Q**-V-	5
FIBRON	1	E-A-**Q**-Q-I-V-Q-P-Q	9
α_2MAC	665	-M-C-P-Q-L-Q-**Q**-Y-E-M-H-G	676
BCAS	162	-V-L-S-L-S-**Q**-S-K-V-L-P-V	173
RHODOP	231	K-E-A-A-A-**Q**-Q-Q-E-S-A-T	242
RHODOP	232	E-A-A-A-Q-**Q**-Q-E-S-A-T	243
RHODOP	233	E-A-A-A-**Q**-Q-Q-E-S-A-T	244
RHODOP	338	S-K-T-E-T-S-**Q**-V-A-P-A-COOH	348

Positively charged residues are in triangles and negatively charged residues are in circles. Note that certain residues such as N-terminal amino acids satisfy both of these charge criteria.

ABBREVIATIONS; FIB=human fibrin; VITRO=human vitronectin; $_2$ P.I.= human α_2-plasmin inhibitor; FIBRON=human fibronectin; α_2-MAC= human α_2-macroglobulin; B-CAS= bovine B-casein; RHOD= bovine rhodopsin.

REFERENCES FOR Table 1.7.: Doolittle et al. (1979) Doolittle et al.(1979) Chen and Doolittle (1971) Skorstengaard et al. (1990) Kimura et al.(1986) McDonagh et al. (1981) Mortensen et al. (1981) Gorman and Folk (1980) McDowell et al. (1986)McDowell et al.(1986) McDowell et al. (1986) McDowell et al. (1986)

1.8.3. UNREACTIVE SITES IN THE PROTEIN SUBSTRATES OF FACTOR XIIIa

Inspection of the amino acid sequences which surround reactive glutaminyl residues in protein substrates of factor XIIIa shows that these sequences include a number of less reactive and unreactive glutaminyl residues. Some of the sequences associated with these less reactive glutaminyl residues are aligned in Table 1.8..

Table 1.8. LESS REACTIVE SITES WITHIN PROTEIN SUBSTRATES OF FACTOR XIIIa

PROTEIN	UNREACTIVE GLUTAMINE	AMINO ACID SEQUENCE
VITRO	(73)	N-A-T-V-H-(E)-Q-V-G-G-P-S-L-T-S
VITRO	(78)	S-L-T-S-(D)-L-Q-A-Q-S-[K]-G-N-P-(E)
VITRO	(80)	T-S-(D)-L-Q-A-Q-S-[K]-G-N-P-(E)-Q-T
*RHODO	(244)	Q-(E)-S-A-T-T-Q-(K)-A-(E)-(K)-(E)-V-T-[R]
*RHODO	(64)	T-L-Y-V-T-V-Q-H-(K)-(K)-L-[R]-T-P-L
*RHODO	(312)	Y-I-M-M-N-(K)-Q-F-[R]-N-C-M-V-T-T
α_2MAC	(870)	M-C-P-Q-L-Q-Q-Y-(E)-M-H-G
CASEIN	(175)	[K]-V-L-P-V-P-Q-[K]-A-V-P-Y-P-Q-[R]

ABBREVIATIONS AND REFERENCES See legend to Table 1.7. Positively charged residues are enclosed in triangles and negatively charged residues are enclosed in circles. *The lysine residues within rhodopsin were succinylated prior to modification of the protein by transglutaminase and therefore these residues carry a negative charge.

The unreactive sites within protein substrates of factor XIIIa resemble sites which are unreactive towards guinea pig liver transglutaminase (see section 1.7.8.4.). This is discussed further in section 1.8.4..

1.8.4. COMPARATIVE STUDIES OF THE SPECIFICITY OF FACTOR XIIIa AND GUINEA PIG LIVER ENZYME

1.8.4.1. INTRODUCTION

The number of reports in which the activity of factor XIIIa and the guinea pig enzymes were simultaneously compared are relatively few : (Brenner and Wold, (1978) ; Gorman and Folk, (1980) ; Gorman and Folk, (1981) ; Chung, (1972). A number of reports are available where separate groups have conducted labelling experiments on the same protein substrates, but using different transglutaminases, eg.

α_2-macroglobulin : GP-liver TGase- (Iwanij, 1977), Factor XIIIa- (Mosher, 1976).

rhodopsin : GP liver TGase- (Pober et al., 1978), Factor XIIIa- (McDowell et al., 1986),

bovine ß-casein: GP liver TGase- (Yan and Wold 1984), Factor XIIIa- (Gorman and Folk, 1980),

1.8.4.2. CONFORMATIONAL DIFFERENCES IN SUBSTRATE SPECIFICITY

On the basis of the results of the studies listed in section 1.8.4.1. it appears that the specificity of guinea pig liver transglutaminase and factor XIIIa is similar towards many of the globular protein substrates tested, although factor XIIIa may be more stringent in its requirements for highly reactive sites. Work with model peptides (Gorman and Folk,

1981) has suggested that differences in specificity between the two enzymes probably derive from features of the linear sequence of amino acids surrounding glutaminyl residues. The short peptides tested for substrate properties are expected to already be highly flexible in solution, although the replacement of individual residues with glycine may further increase local conformational flexibility. Thus it seems more likely that the observed differences in specificity between factor XIIIa and guinea pig liver transglutaminase for short peptides may derive from differences in binding efficiency or catalytic ability associated with the presence or absence of certain amino acid side chains than from conformational factors associated with the formation or loss of gross units of secondary structure.

The enormous improvement of the substrate properties of Gln_{167} for factor XIIIa within β-casein as opposed to the model peptide based on this sequence (Gorman and Folk, 1980), suggests that factor XIIIa may have a requirement for a conformationally strained conformation about this site. If the neighbouring lysine residue Lys_9 contributed to such a favourable conformation in the synthetic peptide, this would also explain the large reduction in substrate properties of the peptide for factor XIIIa following the substitution of Lys_9 for glycine, or following succinylation. The specificity of the liver enzyme for β-casein is much lower than that of factor XIIIa although the specificity for the model peptide based on the Gln_{167} containing sequence was similar to that of factor XIIIa.

1.8.4.3. THE DISTRIBUTION OF CHARGED RESIDUES MAY CONTRIBUTE TO DIFFERENCES IN SPECIFICITY

It appears that positively charged residues associated with the C-terminal side of glutaminyl residues within the flexible regions of protein substrates may reduce the substrate properties for both factor XIIIa (see section 1.8.3.) and guinea pig liver transglutaminase (see section 1.7.) towards flexible non-physiological protein substrates. However both may tolerate the presence of lysine residues at C- terminal positions in proteins which are thought to be physiologically important substrates. Physiological substrates of factor XIIIa which contain positive charge in the region C-terminal to reactive glutaminyl residues include Gln_{84}, Gln_{86} and Gln_{93} of vitronectin and Gln_{328} of fibrin (see section 1.8.1.). This may be due to the evolution of favourable conformations at such sites.

Recent work (Coussons et al., 1991) has provided a means of understanding the specificity of guinea pig liver transglutaminases towards protein substrates on the basis of 1) secondary structure associated with substrate sites and, 2) the distribution of charged, and sterically bulky residues surrounding glutaminyl residues (see chapter 10). It may be that factor XIIIa specificity is determined by a variation on this theme, eg. slightly different patterns of charged residues surrounding glutaminyl sites could differentially effect the substrate properties of a peptide substrate towards the guinea pig liver enzyme and factor XIIIa. In the absence of an unfavourable distribution of charged residues about exposed glutaminyl residues an exposed glutaminyl residue would be predicted to have substrate properties for both enzymes, as has been observed (Gorman and Folk, 1981, 1984).

It may be significant that positively charged residues at positions; -1, and +4 relative to exposed glutaminyl residues within potential protein

substrates, appear to be unfavourable to both enzymes. This may reflect a common structural motif at the active site of both transglutaminases (see chapter 12).

1.9. LESS WELL CHARACTERISED SUBSTRATES OF TRANSGLUTAMINASE(S)

Several proteins have been shown to have substrate properties for guinea pig liver transglutaminase, but the modification site(s) have not been determined. In some cases the tertiary structures of the native proteins are now known and certain glutaminyl residues are strongly implicated as substrate sites purely on the basis of their degree of exposure and the maximum observed stoichiometry of modification eg. rabbit muscle aldolase (see chapter 4). A number of these protein substrates eg. bovine ß-lactoglobulin (see chapter 5), yeast phosphoglycerate kinase (see chapter 6), porcine pepsin (see chapter 8), bovine liver catalase (see chapter 8) have been characterised as transglutaminase substrates. These model substrates have extended the current knowledge of the features which appear to reduce the substrate properties of glutaminyl residues.

In other cases reactive sites must be predicted not only on the basis of exposure, but also on the basis of their lack of features normally associated with unreactive sites. This has been done for a number of proteins including; catalase, and rhodopsin in appendix I of this thesis.

1.10. THE ACTIVITY OF TRANSGLUTAMINASE(S) *IN VIVO*

1.10.1. INTRODUCTION

Tissues in which transglutaminase activity is constitutively expressed are generally exposed to pressure, either mechanical or osmotic in nature. Most have contractile properties and some have bordering functions eg. vascular endothelial cells of the arterial and venous systems, smooth muscle, cardiac muscle etc., (Fesus and Thomazy, 1988). Certain tissues normally have undetectable levels of transglutaminase protein. Little enzymic activity is detected in such tissues unless the enzyme is induced eg. by retinoic acid, (Lichtl et al., 1985), transforming growth factor, (George et al., 1990) or other factors. These tissues act as model systems in which to study the biochemistry of transglutaminase, and in some cases *in vivo* substrates of the enzyme appear to have been identified. The transglutaminase activity associated with some of these tissues is discussed below together with its potential *in vivo* roles.

1.10.2. EXPRESSION OF TRANSGLUTAMINASE ACTIVITY IN LIVER HOMOGENATES

Transglutaminase protein is normally present at low levels in living hepatocytes and is thought to be the same protein as is found in eye lens (see section 1.10.3.). Piacentini and Beninati (1988) have shown that freshly isolated rat hepatocytes contained three times higher levels of putrescine than perfused liver. Exhaustive proteolysis followed by ion exchange and reverse phase HPLC showed that putrescine was the main

attached polyamine, and that spermidine and spermine modified proteins were notably absent.

Russell (1981) has shown that polyamine-protein conjugates may be purified from calf and rat liver several hours after removal of the organ, and that the bulk of this material was located in the 0.6 M NaCl-extracted and chromatin-bound protein fractions of the cells. A major product of transglutaminase modification within liver cells was shown to be ornithine decarboxylase (ODC). Transglutaminase purified from guinea pig liver was shown to incorporate up to 4.0 mol putrescine/mol protein of putrescine into a preparation of ornithine decarboxylase from calf liver *in vitro*. This appeared to cause a loss in activity of ornithine decarboxylase which was directly related to the stoichiometry of amine incorporation into the protein by transglutaminase. Evidence was presented that the ornithine decarboxylase-putrescine conjugate caused a transient increase in activity of RNA polymerase I *in vitro*. Addition of a second aliquot of modified ornithine decarboxylase caused a further increase of RNA polymerase activity, suggesting that the ornithine decarboxylase-conjugate may act as a labile subunit of RNA polymerase I. A model of the regulation of polyamine and RNA synthesis by ornithine decarboxylase and transglutaminase has been presented (Russell 1981). It may be significant that GTP (a precursor of RNA) appears to reduce transglutaminase activity *in vitro* (see section 1.6.7.).

A significant omission in the work presented by Russell (1981) is that the covalent binding of putrescine into ornithine decarboxylase was not directly proven. Unfortunately there is currently not enough structural information concerning ornithine decarboxylase from calf liver to predict which glutaminyl residues within ODC are reactive towards transglutaminase.

However if this protein does prove to be a physiologically important substrate for transglutaminase *in vivo*, then this would suggest a key role for transglutaminase in the control of growth (see section 1.11.4.1.) and differentiation (see section 1.11.4.2.).

Transglutaminase may have a number of physiological roles in liver tissue these include: 1) control of the cell cycle (see section 1.11.4.2.), 2) receptor mediated endocytosis of α_2-macroglobulin (and possibly other protein species) at clathrin coated pits, (see section 1.11.1.), and, 3) a role in apoptosis (see section 1.11.4.3.).

1.10.3. EYE LENS TISSUE

The predominant form of transglutaminase expressed in mammalian eye lens tissue is thought to be identical to that of the liver enzyme. The most effective protein substrates for transglutaminase in bovine eye lens were found to be β-crystallins. The substrate sites, (Berbers et al., 1984) are discussed in section 1.7.8.3.. A gradual accumulation of a dimeric β-Bp-crystallin dimer has been noted in ageing tissue (Lorand et al., 1981). The same product was also formed when guinea pig liver transglutaminase was used to cross-link calf lens homogenates *in vitro* (Lorand et al., 1981). The conservation of substrate sites between species, supports the proposal that the presence of these substrate sites serves some useful role (Berbers et al.,1983).

Lipocortin I is recognised to be the major protein of the EDTA extractable protein of eye lens, Russell et al., (1987). This has been shown to be a substrate in human epidermal carcinoma A 431 cells in the presence of calcium ionophore A231187, (Ando et al., a 1987, b 1991) and from human placenta, from which the reactive glutaminyl site has been determined, (Pepinsky et al., 1989), (see section 1.7.8.3.). Other reported substrates

within this tissue include the cytoskeletal proteins: vimentin, actin (see section 1.7.8.3. for modification site) and both the α- and β- forms of tubulin from both pig brain, and the flagella of *Chlamydomonas reinhardtii* (Iwanij, 1977).

1.10.4. RED BLOOD CELLS

Two types of polymer have been identified in Ca^{2+} activated red blood cells, Lorand (1988): type I, which are membrane associated but extractable with detergents and, type II which remain with the alkali stripped matrix of the ghost.

The determination of the components of each type of complex has so far identified ankyrin (band 2.1), both α-, and β-spectrins, band 4.1 (see section 1.4), catalase (see section 7.2) and haemoglobin (see section 1.7.7.1.) in both types of complex. They appeared to be attached to band 3 (anion transporter protein) in type II complexes. The stoichiometry of modification of some of these proteins have been estimated, (Dutton and Singer, 1975).

1.10.5. EPIDERMAL TISSUE

During the late stages of terminal differentiation of epidermal keratinocytes a highly insoluble matrix of cross-linked protein becomes deposited beneath the plasma membrane to form a cornified envelope, (Green, 1979). Transglutaminase(s) are thought to stabilize this structure through the formation of Gln-Lys cross-links, (Rice and Green, 1977). In epithelial cells the cytokeratins are the intermediate filaments that form

the cytoskeleton. This creates an appealing link between the physiological roles of sequence-related transglutaminases in keratinocytes and in cirrhotic hepatocytes [(see section 1.11.4.4. (c)]. In both tissues transglutaminases cross-link components of the cytoskeleton during the process of apoptosis (or programmed cell death), (see section 1.11.4.3). The best characterized substrate proteins within the epidermal cell envelope include involucrin, (Simon and Green, 1988), and loricrin, (Mehrel et al., 1990). The known modification sites within involucrin are aligned in Fig. 1.17.†.

491 Q-L-E-L-P-E-Q-Q*-V-G-Q-P-K- 504

566 L-L-P-V-E-H-Q-Q*-Q-K-Q-K-V- 579

Fig. 1.17. SUBSTRATE SITES FOR TRANSGLUTAMINASE IN INVOLUCRIN

†: It should be noted that the epidermal transglutaminases may have different specificity towards protein substrates than the enzyme isolated from guinea pig liver. The reactive glutaminyl residue is indicated with an asterix*

The most reactive glutaminyl sites for the keratinocyte transglutaminase are characterized by the presence of glutamic acid residues in the region N-terminal to reactive glutaminyl residues.

The primary structure of a second cell envelope protein has recently been determined from human keratinocytes (Mehrel et al., 1990). This protein also appears to be a substrate for an intracellular transglutaminase and the position of Lys-Gln cross-links has been determined (Hohl et al., 1991). The substrate sites within loricrin are shown in Fig 1.18..

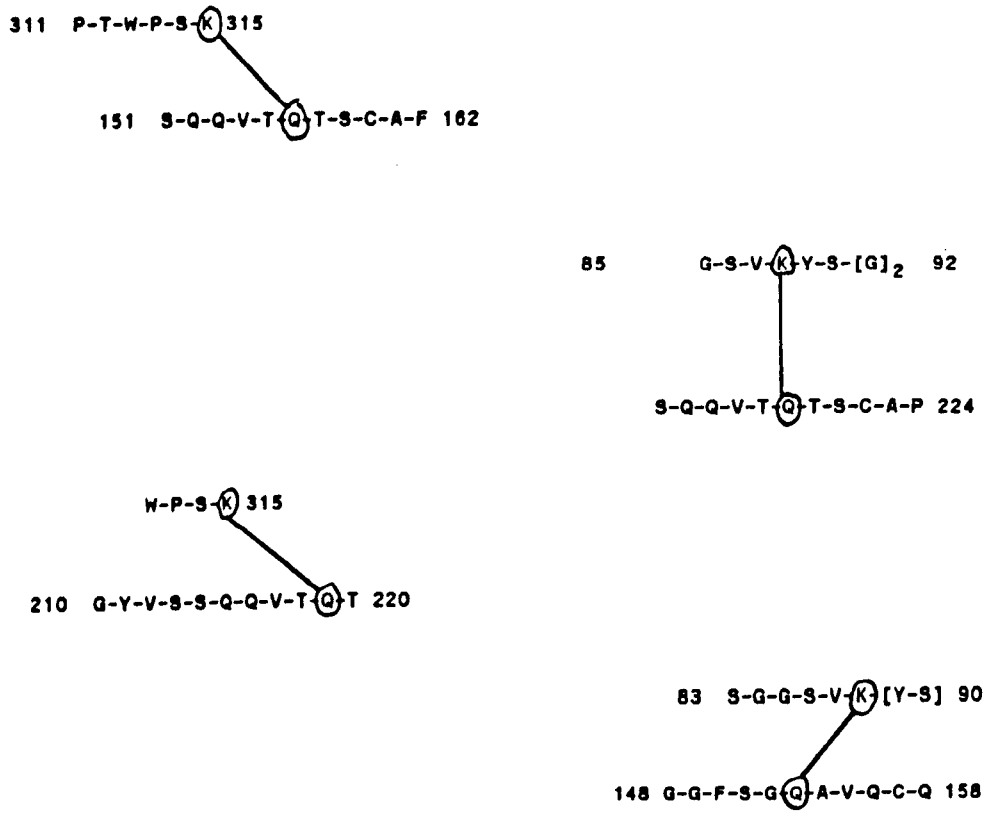


Fig 1.18. SUBSTRATE SITES FOR TRANSGLUTAMINASE WITHIN HUMAN LORICRIN O—O=ISOPEPTIDE BOND

The substrate sites in loricrin for the keratinocyte transglutaminase resemble those within the substrates of the guinea pig liver enzyme (see section 1.7.) in: 1) their content of hydroxylated and small hydrophobic amino acids, 2) the absence of positively charged residues in the C-terminal regions, and 3) low proportion of large sterically obstructive side chains eg. Tryptophan, phenylalanine, isoleucine, leucine etc.

1.11. METABOLIC ROLES FOR TISSUE TRANSGLUTAMINASES

1.11.1. RECEPTOR MEDIATED ENDOCYTOSIS

Primary amines and glutaminyl peptides which have the common property of being competitive inhibitors of transglutaminases, can inhibit the internalization of ligands through coated pits and vesicles (Davies et al., 1980). Some of the proteins which appear to be internalized into cells by this route have been identified as substrates of transglutaminase *in vitro*. These include α_2-macroglobulin, [for which a substrate site for factor XIIIa has been determined *in vitro*, (see section 1.8.)]. This phenomenon has been reviewed by Fesus et al., (1984).

1.11.2. STIMULUS-SECRETION COUPLING

Transglutaminase activity has been observed in the pancreatic islets of Langerhans (Bungay et al., 1986). The secretion of insulin from these cells may be inhibited using competitive inhibitors of transglutaminase. It has also been shown that in parallel with insulin secretion, high Mr protein polymers which are insoluble in SDS are formed (see section

1.11.3.). The transglutaminase catalysed incorporation of [^{3}H] histamine into a Mr 84000 particulate protein has been noted (Gomis et al., 1989), although the presence of Glu-Lys cross-links has not been determined in this material.

1.11.3. FORMATION OF INTRACELLULAR MATRIX

The existence of a urea-, guanidine HCl-, detergent-, and mercaptoethanol- insoluble intracellular matrix which is cross-linked may form the basis of a very fine filament structure in cells (Loewy et al., 1983). Similar material has been obtained from a variety of cells and tissues even in the absence of previous stimulation or manipulation. In no case have Glu-Lys bonds been located in such materials and hence the involvement of transglutaminases in the production of these materials remains speculative. The protein composition of this material has not yet been determined. though preliminary data suggests that it may contain titin (Fesus and Thomazy 1988).

1.11.4. CELL-CYCLE REGULATION

1.11.4.1. CELL GROWTH

The activity of intracellular transglutaminases during cell growth is difficult to measure since most systems consist of mixtures of cells at different stages of development. However polyamine biosynthesis is known to be related to cell growth (Tabor and Tabor, 1984) and since these compounds may act as substrates for transglutaminases, some workers

have investigated the changes in transglutaminase activity which are known to accompany perturbation of the levels of primary amines in cells.

Following the addition of [^{3}H] putrescine to growing liver cells it has been shown that putrescine may become covalently incorporated into cellular proteins (Russell, 1981) [see section 1.10.2.]. A similar result was obtained when transglutaminase activity was measured during the early development of sea urchin cells (Cariello et al., 1984). A 30 kDa protein was observed to be particularly susceptible to modification *in vivo* by spermidine (Canellakis et al., 1985). Six hours after fertilization of sea urchin eggs, approximately 7 % of the total cell associated spermidine appeared to be detected bound in a macromolecular form. In this (as in other cases) it is difficult to dissect out the effects which are due to differentiation and those due to growth since these processes are both required during development.

1.11.4.2. CELL DIFFERENTIATION (CONTROL OF CELL CYCLE)

Marked variation in the levels of cellular polyamine content has been observed in association with cell differentiation (Tabor and Tabor, 1984), though the role exerted by these compounds during the expression of the differentiated phenotype is still poorly understood. An increase in levels of these compounds might be expected to:

1) reduce the ability of transglutaminase to cross-link proteins by Glu-Lys bonds, hence destabilizing structures within the cell.

2) increase the levels of polyamine derivatives of physiologically important protein substrates, thus facilitating membrane attachment and/or controlling turnover rates.

The modification of cytoskeletal proteins by transglutaminase appears to increase the rate and extent of microtubule assembly *in vitro* in the case of both, tubulin (Maccioni and Seeds, 1986) and actin (Takashi, 1988). These findings support the notion of a role for transglutaminase in the differentiation of cells *in vivo*. In order to investigate the role(s) of transglutaminase in the control of the cell-cycle some workers have considered the effects on transglutaminase activity and expression following the breakdown of the cell cycle in tumorigenic tissues [see section 1.11.4.4. (a)]. Evidence for a controlling role for transglutaminase in normal liver cells is discussed in section 1.10.2..

Investigation into the incorporation of polyamines into the cell body proteins of the giant axon of *Aplysia* showed that tubulin was labelled (Ambron et al., 1982). Within mouse neuroblastoma cells both actin and tubulin were found to be labelled (Maccioni and Seeds, 1986). Label was also incorporated into; 92 kDa, 76 kDa, and 72 kDa proteins on the surface of differentiated neuroblastoma cells though not in undifferentiated counterparts (Chen, 1984).

1.11.4.3. CELL DEATH (APOPTOSIS)

In many tissues, transglutaminase activity is characterized by the formation of high Mr proteinaceous complexes which cannot be dissociated by treatment with SDS and DTT (see section 1.11.3.). This phenomenon may be inhibited by the feeding of primary amines to such cells, and has been

termed apoptosis or " programmed cell death". The red blood cell has been considered an *in vitro* model system in which to study this phenomenon and has been reviewed by Lorand (1988), (see section 1.10.4.).

Keratinocytes are epithelial cells which appear to have evolved specifically to undergo apoptosis in order to produce the protective stratus corneum of skin. It is possible that the induction of transglutaminase activity associated with the terminal differentiation of keratinocytes may also occur in a variety of other cell types (possibly under different control mechanisms), and that in each case the effect is destruction of the cell.

Apoptosis may be thought of as a means of regulating cell proliferation. Transglutaminase-catalysed cross-links appear to be maintained at a minimal level under proliferating conditions. However when cells cease proliferating, transglutaminase activity increases and cells may be stabilized by the increased levels of cross-linked proteins. Section 1.11.4.2. considers the opposite series of events associated with differentiation.

The levels of free calcium ions normally associated with the cytoplasm of liver cells is estimated to be in the 10^{-7} to 10^{-5} M range (Klee et al., 1980). This is not likely to be sufficient to activate transglutaminase (see section 1.6.2.). Hence if apoptosis is regulated by transglutaminase the levels of calcium would probably have to be substantially increased. The disintegration of mitochondria (which act as intracellular Ca^{2+} stores) has been suggested as the cause of activation of the intracellular epidermal transglutaminase *in vivo* (Rice and Green, 1979). Apoptosis may be under much finer control than suggested by this simple model.

1.11.4.4. CELL-CYCLE BREAKDOWN

a) TUMORIGENESIS

A reduction in transglutaminase activity has been noted in viral and chemically transformed cells (Birckbichler et al., 1976, 1977). Investigation of transglutaminase activity in hepatocellular carcinoma indicated;

1) increased levels of incorporation of radioactive amines into the soluble fraction of cell proteins.

2) decreased levels of Gln - Lys cross links compared to normal liver, (Knight et al., 1990).

The presence of high levels of isopeptide bonds has previously been associated with the non-proliferating state, (Birkbitcher et al., 1978). Labelling of tissue slices indicates that the majority of [^{14}C]-putrescine is incorporated into the high Mr fraction associated with the plasma membrane (85%) and that the level of incorporation into this region is comparable in both tumorigenic and in normal liver cells. It appears that the particulate transglutaminase is more active in terms of incorporation of amines into proteins, while the cytosolic enzyme is more active in cross-linking proteins by Glu-Lys bonds.

A comparison of the distribution of labelled proteins between normal and transformed tissues revealed some differences, (Hand et al., 1990). Proteins of Mr 39,000 and 44,000 were labelled in both cancerous and normal tissue, although the levels were reduced in the tumorigenic tissue. In addition the tumour tissue exhibited additional radiolabelling in a protein of Mr 36,000.

An inactive 120 kDa transglutaminase variant has been located in rat fibrosarcomas and this correlated with the reduction in the levels of the cytosolic form. Restoration of activity to the 120 kDa form was achieved by partial proteolysis by trypsin and chymotrypsin, (Knight et al., 1990). Hence it appears that in metastatic tumour cells, the reduction in the levels of cytosolic transglutaminase activity is a consequence of its perturbed expression possibly as a result of inappropriate post-translational processing or inappropriate gene expression.

The activity of cytosolic keratinocyte transglutaminase was increased by a factor of ten times in normal cell cultures relative to transformed keratinocytes by exposure to low doses of transforming growth factor ß-1, (George et al., 1990). This appeared to be due to both increased transcription and translation. The lack of responsiveness of the transformed cells was found to not be due to lack of receptors but to some difference in signal transmission. It should be noted that transglutaminases are capable of GTP hydrolysis and hence may have some ability to act as a G-proteins (see section 1.6.7.) *in vivo*.

b) IN ALZHEIMERS DISEASE

A transglutaminase activity has been detected in human brain tissue, (Selkoe et al., 1982). Similarities have been observed between the neurofibrillary tangles associated with sufferers of Alzheimers disease and the cross-linked products of the action of transglutaminase on neurofilament enriched preparations purified from white matter. Protein substrates included : myelin basic protein (20 kDa), together with a 200 kDa, 160 kDa (and less clearly) a 68 kDa protein as assessed by SDS-PAGE. Selkoe et al., (1982) have postulated a role for transglutaminase in the

aetiology of the disease. It may be more than coincidence that antibodies to the ß-amyloid peptide cross-react with conformational epitopes within fibrinogen (Stern et al., 1990) which is an efficient substrate for guinea pig liver transglutaminase (Chung, 1972).

c) CIRRHOSIS OF THE LIVER

It has been shown that griseofulvin treatment (which disrupts cytoskeletal structure) considerably increases transglutaminase activity. Zatloukal et al., (1989) have shown that hepatocellular cytokeratins may act as substrates for transglutaminases *in vitro* to form highly insoluble lattices which may be the components of Mallory bodies in the livers of patients with acute alcoholic hepatitis. This may be analogous to cytokeratin cross-linking during the terminal differentiation (apoptosis) of keratinocytes, see section 1.11.4.3..

1.12. THE AIMS OF THIS PROJECT

1.12.1. DETERMINATION OF SPECIFICITY OF GUINEA PIG LIVER TRANSGLUTAMINASE

The primary aim of this project was to investigate the features within polypeptide and protein substrates which determine the specificity of the cytosolic transglutaminase from guinea pig liver towards its substrates *in vitro*. This information could then be used to determine the feasibility of engineering reactive glutaminyl residues into protein substrates.

1.12.2. TO DETERMINE NOVEL USES FOR TRANSGLUTAMINASE

From the information presented in this introduction it is clear that transglutaminases are useful to the protein chemist as a means of cross-linking a wide variety of biological and non-biological materials under the mild reaction conditions of an enzyme catalysed reaction. Since many of the known labelling sites within proteins are in surface exposed, flexible regions, transglutaminases may be used to;

1) Identify regions of secondary structure in proteins. The attachment of reporter molecules to the surface of proteins may also assist in the determination of the proximity of other groups within the structure. This approach has been used to probe the structure of bovine rhodopsin (Pober et al., 1978). A similar approach could be used to investigate the structural basis of conformational change in proteins associated with partial denaturation eg. that seen in bovine β-casein (see chapter 7) at 20 mM Ca^{2+}, (Cooke et al., 1974).

2) Transglutaminases could be used to confirm the identity of glutaminyl residues within peptides of uncertain sequence.

3) Reactive glutaminyl sites could be useful for the attachment of pharmacologically useful amines on to antibody molecules for use in clinical chemotherapy. The feasibility of constructing such conjugated products is discussed in chapter 13.

1.12.3. DETERMINATION OF THE ROLE(S) OF LIVER TRANSGLUTAMINASE

Although this introduction contains considerable speculation concerning the physiological role(s) of the liver transglutaminase, none of these roles are yet firmly established. While it was not a direct aim of this project to determine the biological role(s) of the enzyme it is clear that the study of the basis of the specificity of transglutaminase *in vitro* may give clues concerning the nature of protein substrates *in vivo*. An attempt has been made to address this problem by the use of computer based techniques see section 9.3.2..

CHAPTER TWO: MATERIALS AND METHODS

2.1. MATERIALS

2.1.1. PROTEINS AND PEPTIDES

Proteins and peptides were obtained from the following sources; Transglutaminase (guinea-pig liver), tosylphenylchloromethane (TPCK) treated trypsin and tosyllysylchloromethane (TLCK) treated chymotrypsin from bovine pancreas, V8 proteinase from *Staphylococcus aureus*, catalase from bovine liver, pepsin from porcine stomach, ß-casein and ß-lactoglobulin from bovine milk, cytochrome C from horse heart, ribonuclease A from bovine pancreas, concanavalin A from jack bean, bovine insulin A chain, ovalbumin and lysozyme from chicken egg white, and Dalton mark VII SDS-PAGE Mr marker set, were supplied as lyophilized powders from Sigma Chemical Co, Poole Dorset U.K.. Alcohol dehydrogenase from yeast, aldolase from rabbit muscle, malate dehydrogenase from porcine heart, 3-phosphoglycerate kinase from yeast, glyceraldehyde-3-phosphate dehydrogenase from yeast were supplied as ammonium sulphate suspensions by Boehringer BCL Ltd, Lewes, East sussex. Wild type PGK (yeast) and a number of mutant forms including: His_{388} Gln_{388}, Arg_{168} Met_{168}, and Arg_{168} Lys_{168} forms, were prepared as described by Perkins *et al.* (1983) and Wilson *et al.*, (1987). Plasmids pMA27 and pMA40b were kindly given by Dr. L. Gilmore, Department of Biochemistry, University of Edinburgh, Edinburgh EH8 9XD, Scotland, U.K. The proteins were purified by Dr C. Johnston, Department of Chemistry, University of Glasgow, G12 8QQ, Scotland. Dehydroquinase from *E. coli* was a gift of Dr. C. Kleanthous, Department of Biochemistry, Glasgow University, Glasgow. G128QQ.

A synthetic peptide, shown in Fig 2.1., is based on the known (factor XIIIa catalysed cross-linking site of human fibrin (Chen and Doolittle, 1971);

NH_2-L-T-I-G-E-G-Q-Q-H-H-L-G-G-A-K-Q-A-G-D-V-COOH

Fig 2.1. THE SEQUENCE OF A SYNTHETIC PEPTIDE BASED ON THE HUMAN FIBRIN

This peptide was synthesized by Dr Chris Bose of Celltech Ltd., Slough, Berkshire. SL14DY. Celltech Ltd. also provided samples of IgG. B72.3 (an antibody raised against mucein), a metalloproteinase inhibitor (TIMP). The synthetic peptide CBZ-Gln-Gly, was from Cambridge research biochemicals Ltd., Cambridge, England, CB25nX.

2.1.2. CHEMICALS

2.1.2.1. GENERAL LAB CHEMICALS

Guanidine hydrochloride (Ultrapure grade) was from BRL Betheseda Research laboratories, Life technologies Ltd., Gaithersburg, MD 20877 U.S.A. Acrylamide, and N,N'-methylene bis-acrylamide (electran grade), EDTA [ethylenediaminetetraacetic acid (dinatrium salt)] were from BDH Ltd. Ammonium sulphate (especially low in heavy metals), and Calcium Chloride (6.0 H_2O) were from Fisons Ltd. DTT was from Sigma. Optiphase "M.P." scintillation fluid was from FSA Lab supplies, Loughborough. SDS, Trizma

base (Tris[hydroxymethyl]amino-methane), trichloroacetic acid, triethanolamine, MOPS (3-[N-morpholino]propane sulphonic acid), TEMED, (N,N,N',N'tetramethylethylenediamine), 2-Mercaptoethanol, ammonium bicarbonate, monodansylcadaverine hydrochloride, putrescine dihydrochloride, were from Sigma Chemical Co. Substrates: ADP, ATP, NAD+, NADP+, NADH, NADPH were supplied as lyophilized powders from Boehringer BCL Ltd.

2.1.2.2. RADIOCHEMICALS

1,4-[^{14}C-]Putrescine dihydrochloride (4.07GBq/mmol) was from Amersham International plc. Amersham, Bucks, U.K.

2.1.2.3. ORGANIC SOLVENTS

Propan-2-ol and acetonitrile (Far u.v. HPLC grade) were from Labscan Ltd., Dublin, Eire. Trifluoroacetic acid was from Rathburn Chemical Co., Walkerburn, Peebles, Scotland, U.K. Triethylamine (Hipersolv) was from BDH Ltd.

2.1.2.4. STORAGE OF CHEMICALS

Ammonium sulphate suspensions were stored in sealed glass containers at 4°C. Lyophilized powders were stored at -20°C. Solvents were stored in the dark at room temperature.

2.1.3. CHROMATOGRAPHY MATERIALS

2.1.3.1. HPLC COLUMNS

Vydac C-18 reverse phase end capped reverse phase h.p.l.c. column, (dimensions: 25 cm x 4.6 mm) particle size 5 µm; pore diameter 300 Å, was from Technicol, Ltd., Stockport, Cheshire, U.K.. Zorbax Bio series C-8 PEP RP.1 reverse phase h.p.l.c. column (dimensions: 8 cm x 6.2 mm) pore size 5 µm was from Dupont (U.K.) Ltd, Wedgewood Way, Stevenage, Herts SG1 4QN.

2.1.3.2. GEL-FILTRATION/ION EXCHANGE MEDIA AND COLUMNS

Sephacryl S-300 and disposable P.D.10 (Sephadex G25M) columns were purchased from Pharmacia Ltd., Milton Keynes, Bucks., U.K.. DEAE sephadex was purchased from Sigma Chemical Co.

2.1.4. MISCELLANEOUS ITEMS

Cellulose filter discs (3 MM : 21 cm diameter) were from Whatman International Ltd., Maidstone, Kent., U.K.

2.2. METHODS

2.2.1. INTRODUCTION

A variety of different approaches have been used in order to investigate the specificity of transglutaminase (see chapter 1). An ideal approach would be capable of determining the effects of small single changes in conformational or chemical features surrounding a single reactive glutaminyl residue on the substrate properties of a protein of known 3-D structure. Unfortunately such a precise approach requires equipment and materials which were not available for the purposes of this project. In lieu of this, the approach calculated to provide the most information concerning the specificity of transglutaminase was considered to be one which would;

1) increase knowledge concerning the reactivity of transglutaminase towards conformationally constrained model substrates (ie. proteins of known sequence and 3-D structure).

2) facilitate pattern searching within the known and novel substrate sites.

The substrate properties of a number of structurally diverse proteins were investigated and then compared and contrasted with the substrate sites published prior to the project (see chapter 1). Sections 2.2.2. to 2.2.8.4. give details of the general procedures used for this investigation. Any differences in these procedures are detailed separately in the text describing the results of this work.

2.2.2. GENERAL LABORATORY METHODS

2.2.2.1. PREPARATION OF PROTEIN AND PEPTIDE SUBSTRATES OF TRANSGLUTAMINASE

Transglutaminase is rapidly inactivated by low (<10 mM) concentrations of ammonium sulphate and thus this was removed from proteins prior to testing their substrate properties. Proteins were routinely desalted by dialysis [see section 2.2.2.3. (a)].

Proteins and peptides which were supplied as lyophilized powders were allowed to slowly warm up to room temperature before opening. Buffer (also 15 °C) was added and the protein was assisted to dissolve when necessary by gentle swirling (foaming was carefully avoided). Proteins solutions prepared in this way were incubated for a minimum of an hour at 15°C prior to inclusion into reaction mixtures. Any insoluble aggregates were removed by centrifugation at 1500 rpm for 2 minutes using a Microcentaur bench centrifuge.

2.2.2.2. THE DETERMINATION OF PROTEIN CONCENTRATIONS

Protein concentration was routinely determined during transglutaminase purification using the method of Sedmak and Grossberg, (1977). Protein samples were made up to 1.0 ml with distilled water. To this was added 1.0 ml of Coomassie reagent which consisted of 0.06 % wt/vol Coomassie Brilliant Blue G-250 in 3 % vol/vol perchloric acid. The absorbance of the resulting solution was measured against an absorbance blank which had water substituted for protein solution. The ratios of the absorbance of the solutions at 620 nm/465 nm (corrected for the blank), were calculated and the protein concentrations calculated relative to a standard curve

prepared using known concentrations of bovine serum albumin (BSA) in the 0-50 µg/ml range. A new standard curve was constructed for each new batch of Coomassie reagent.

For purified proteins the concentrations were calculated on the basis of published absorption coefficients these are given when required in the text.

2.2.2.3. DESALTING OF PROTEINS

a) DIALYSIS

Protein samples were pipetted into pre-boiled dialysis sacks (exclusion limit ~ 8 kDa) and were dialysed with stirring at 4°C (in the dark in the case of dansylated protein) against repeated changes of buffer. A typical buffer used was triethanolamine hydrochloride (50 mM) adjusted to pH 7.6 with 1 M NaOH. Typically, between 1.0 and 3.0 ml of protein or peptide suspension was first dialysed against five litres of buffer at 4°C for four hours, with stirring. The buffer was then changed and a second dialysis continued overnight.

b) SIZE EXCLUSION CHROMATOGRAPHY

Dansylcadaverine and putrescine were routinely removed from proteins using commercially available G-25 Sephadex columns (dimensions 6 cm x 1.5 cm). Typically between 0.5 and 1.5 ml of protein solutions were loaded at concentrations of between 1-5 mg/ml. They were then eluted with a 1% w/v solution of ammonium bicarbonate pH 8.0.

2.2.2.4. PREPARATION OF PRIMARY AMINE SUBSTRATES.

a) PREPARATION OF DANSYLCADAVERINE STOCK SOLUTIONS

Dansylcadaverine hydrochloride is only sparingly soluble in aqueous solution at neutral pH, and was prepared as a stock solution using the method of Lorand and Campbell, (1971), which is as follows:

1) Dansylcadaverine dihydrochloride (20 mg) was weighed into a glass bottle (20 ml capacity) and 10 ml of an appropriate buffered solution was added e.g. Tris-HCl pH 7.6 (100 mM).

2) Dilute hydrochloric acid was added drop-wise to reduce the pH of solution to 2.0, whereupon the crystals dissolve and the solution decolorises. The pH of solution was then raised to pH 7.6 by the drop-wise addition of NaOH (1.0 M) (aq) with continuous stirring.

3) The solution was then left to settle at 4°C for 1 hr, before it was filtered through a Millepore filtration cassette (pore size 0.22 μm) under pressure exerted by use of a syringe. The filtrate clarifies to a pale yellow/green solution. The concentration of the stock solution may then be adjusted by dilution with Tris buffer (100 mM) pH 7.6.

In order to quantify dansylcadaverine, fluorescence measurements were made at 20°C in a Perkin - Elmer MPF-3L fluorometer, using semi-micro quartz cuvettes of 1.0 ml capacity. A calibration curve was constructed using an excitation wavelength of 320 nm (slit width 4.0 nm). The emission wavelength was measured at 520 nm (slit width 8.0 nm). The concentration of dansylcadaverine was calculated from the absorbance of appropriately diluted 1 ml samples at 330 nm using an $LKB_{BIOCHEM}$ 4050 Ultraspec II uv/vis spectrophotometer.

The fluorescence of samples was calculated in fluorescence units/mmol and were checked regularly as the fluorescence of dansylcadaverine was found to reduce with time in solution (possibly due to photochemical damage). The absorbance of stock solutions at 330 nm remained constant [A 1.0 mM solution of dansylcadaverine has an A_{330} of 4.65]. The absorption spectrum is shown in Fig 2.2. A 10 mM stock solution of dansylcadaverine is stable in solution at 4°C for several months if kept in the dark.

b) PREPARATION OF [^{14}C]-PUTRESCINE STOCK SOLUTIONS

Putrescine dihydrochloride is highly soluble in all the buffered solutions described in this thesis and was dissolved prior to use to give a clear colourless solution. Typically a 10mM [1.61 mg/ml] stock solution was prepared.

Radioactive stock solutions were prepared by dilution of a 0.4 mM stock solution of [^{14}C]-putrescine (50 µCi/ml) with the stock (10 mM) solution of non-radioactive putrescine to give a radioactive stock solution of 9 mM (4.097 GBq.mmol). This was subsequently diluted into incubation mixtures as described in section 2.2.3.2.. Stock solutions of putrescine were stored at -20°C when not in use.

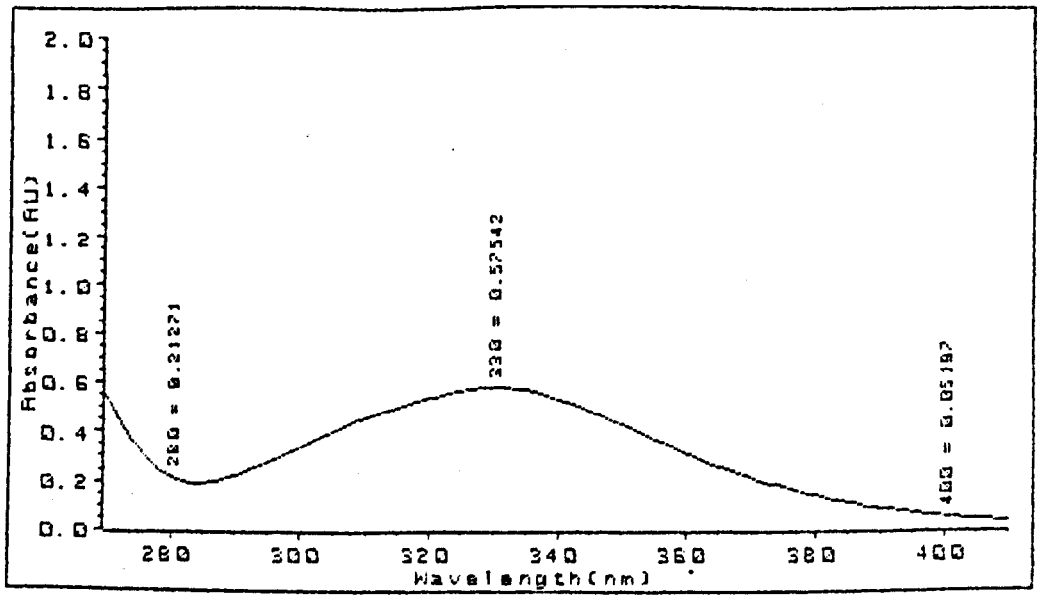


Fig 2.1. THE ABSORBTION SPECTRA OF DANSYLCADAVERINE

A stock solution of 12.0 mM of DNSC was diluted 100 fold with tris-HCl (50 mM) pH 7.6 and 1 cm^3 of this was analysed over a wavelength range from 270 to 400 nm. The A_{330} of a 1 mM solution of DNSC is therefore ~4.65.

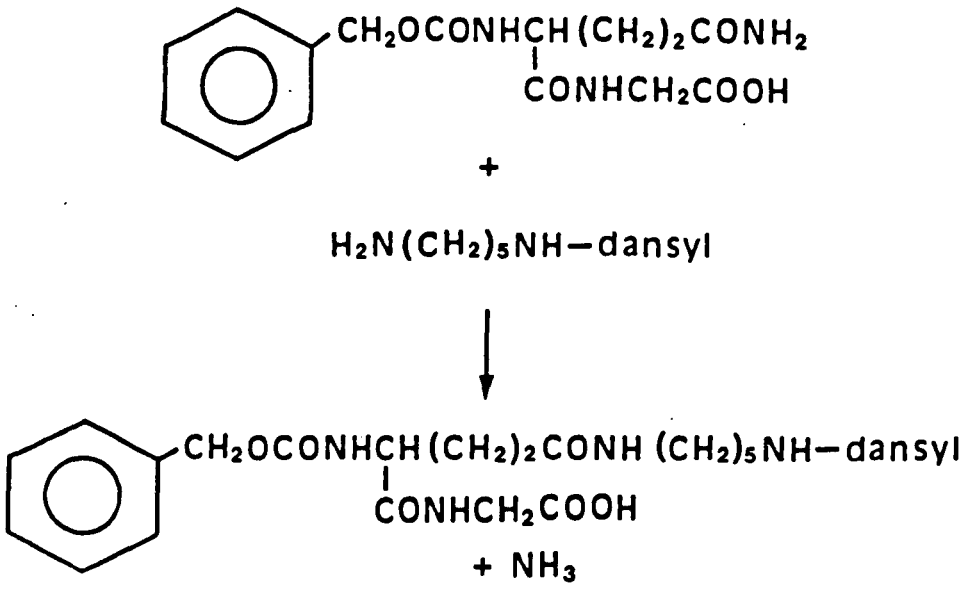


Fig 2.2. THE COVALENT STRUCTURE OF THE SYNTHETIC PEPTIDE CBZ-GLN-GLY

CBZ-Gln-Gly is shown prior to reaction in (a) and following transamination with the amine group of dansylcadaverine in (b). The covalent structure of dansyl cadaverine is shown in greater detail in fig 2.3. In the colorimetric reaction hydroxylamine replaces dansylcadaverine as the amine substrate and the product quantitatively forms a red solution with acidified iron salts.

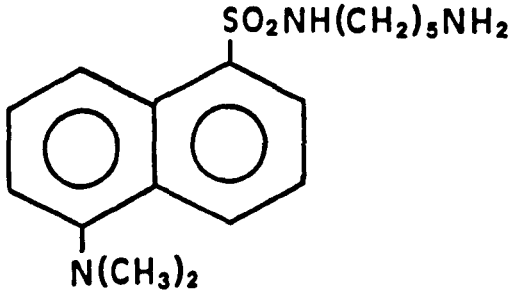


Fig 2.3. THE COVALENT STRUCTURE OF DANSYLCADAVERINE (DNSC)

Systematic name is N-(5-Aminopentyl)-5-dimethyl-amino-1-napthalenesulphonamide.

2.2.3. THE ASSAY OF TRANSGLUTAMINASE ACTIVITY

2.2.3.1. A COLORIMETRIC ASSAY FOR TRANSGLUTAMINASE ACTIVITY

The synthetic peptide CBZ-Gln-Gly (Fig 2.3) was used as a substrate for the routine assay of transglutaminase using the method of (Folk and Cole, 1966). The incorporation of hydroxylamine results in the formation of a hydroxamate derivative. This is stable at acidic pH and will quantitatively form a purplish-brown ferric hydroxamate chelate with an iron III chloride/TCA mixture, (Lipmann and Tuttle, 1945). The formation of this complex may be monitored from the increase in absorbance at 520 nm. A standard curve was constructed using known amounts of hydroxamic acid which has an identical molar extinction coefficient to the hydroxamic acid derivative of CBZ-Gln-Gly after complex formation with acidified iron III chloride (Lippman and Tuttle, 1945). This assay procedure was used to follow the purification of transglutaminase (see chapter 3).

Table 2.1. The ingredients for the colorimetric assay of transglutaminase

compound	stock concentration (mM)*	volume added (ul)	final concentration (mM)
CBZ-Gln-Gly (aq)	30.0	115	10.0
NH2OH (aq)	200.0	175	100.0
Ca2+ (aq)	100.0	20	6.0
DTT (aq)	50.0	10	1.5
Tgase (aq)	1 mg/ml	25	(0.07mg/ml)

* except transglutaminase (Tgase)

PROCEDURE

1) Add ingredients to assay mixture as shown in Table 2.1. [all dissolved in Tris-HCl pH 7.0 (200 mM)], in the order: CBZ-Gln-Gly, NH_2OH, DTT, Ca^{2+}, then transglutaminase.

2) Incubate at 37°C for 1hr.

3) Stop reaction by addition of 20 µl of EDTA (37 mM) to a final concentration of 5mM.

4) Add 0.5 ml of freshly prepared acidified iron III chloride (Soln X)*.

5) Spin at 2,000 rpm (to remove denatured protein) for three minutes until clear.

6) Measure absorbance of solution in a 1 cm plastic cuvette at 520 nm and subtract blank (reaction mixture minus enzyme subjected to the same procedure). The rate of CBZ-Gln-Gly-hydroxamate derivative formation may then be calculated relative to a standard curve (Fig 2.4).

* PREPARATION OF SOLUTION X

I) The following stock solutions were prepared;

a) $FeCl_3$ (5 % w/v) in 0.1 M HCl, 1 vol.

b) 3M HCl 1 vol

c) TCA (12 % w/v) in H_2O.

II) The components were mixed in a 1vol/1vol/1vol ratio.

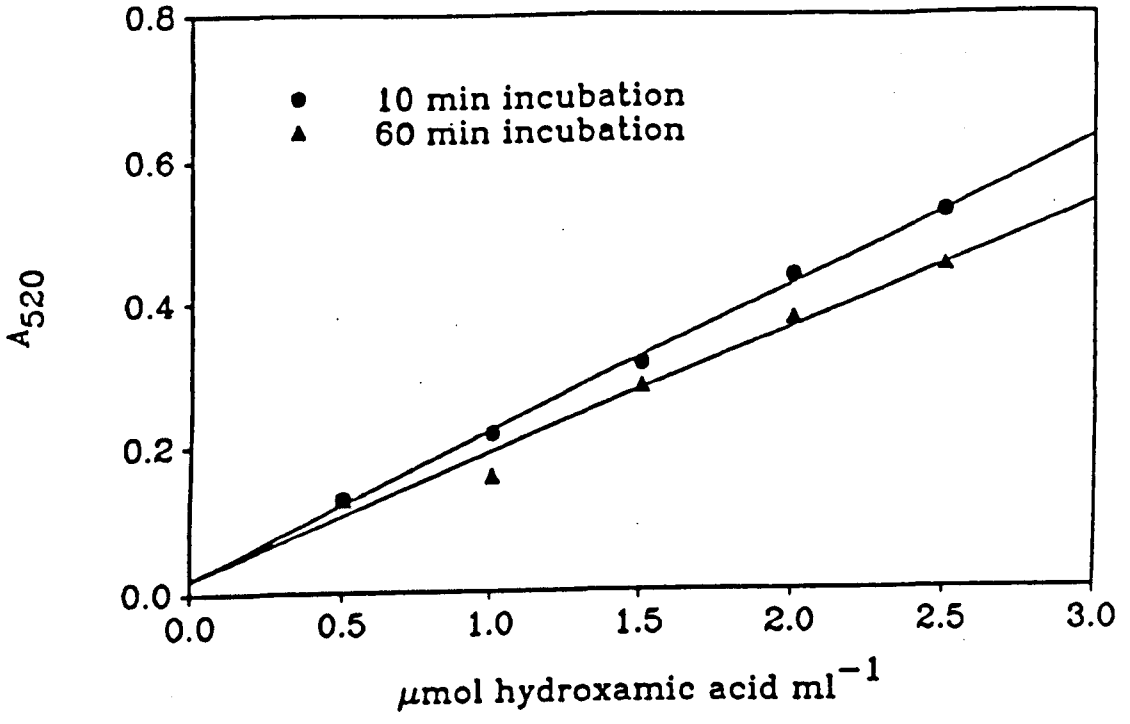


Fig 2.4. A STANDARD CURVE FOR THE ESTIMATION OF THE PRODUCTS OF THE COLORIMETRIC ASSAY OF TRANSGLUTAMINASE.

Since guinea pig liver transglutaminase has been shown to form γ-glutamohydroxamic acid derivatives from peptide substrates using hydroxylamine as the acyl acceptor this may form the basis of a colorimetric assay since these products of reaction can quantitatively form coloured complexes with ferric salts. The compound γ-glutamohydroxamic acid has been shown to develop similar colour equivalent and has a similar absorption spectra to the glutamohydroxamic acid derivative of CBZ-Gln-Gly (Grossowicz et al., (1950) and hence a standard curve for the production of this compound may be constructed by incubating known amounts of glutamohydroxamic acid with iron salts and monitoring the colour development. Stock solution (0.4 mg/ml) of glutamohydroxamic acid was prepared in the standard incubation mixture used for the colorimetric (CBZ-Gln-Gly) assay of transglutaminase. To 2 ml of this, 2.0 ml of acidified ferric chloride solution was added. The mixture was incubated for 10 min at 37°C and then the solution was diluted with water to the required concentration for analysis. The O.D. was read at 520 nm after 10 min, and 60 min.

2.2.3.2. A RADIOLABEL-INCORPORATION ASSAY OF TRANSGLUTAMINASE ACTIVITY

Transglutaminase activity was determined by the determination of the rate of enzyme-catalysed covalent incorporation of [^{14}C]-putrescine into bovine ß-casein, using a method based on that of Dvilansky et al., (1970). The components of the reaction mixture (shown in Table 2.2.) were combined, and samples removed at appropriate time intervals. Samples were pipetted onto Whatmann 3MM cellulose discs, and the modified protein was precipitated by immersion of the disc(s) in a bath of cold TCA. Any unbound radiolabelled putrescine may then be removed by repeated washing of the discs in hot trichloroacetic acid (TCA), by the method of Mans and Novelli, (1961). The procedure is described below.

Table 2.2. The ingredients used for the radioactive assay of transglutaminase activity

compound (mg/ml)	stock concentration (mM)*	volume added (ul)	final concentration
ß-casein (aq)	(1.0 - 10.0 mg/ml)	150	(0.43-4.3 mg/ml)
[14C] putrescine	10.0	150	4.3
CaCl2 (aq)	100.0	15	4.3
DTT (aq)	50.0	6	1.2
transglutaminase	1 mg/ml	25	(0.07 mg/ml)

* except proteins

PROCEDURE

1) Add the components listed in Table 2.2. to the assay mixture in the order: bovine ß-casein, [^{14}C]-putrescine, Ca^{2+}, DTT, and finally enzyme. The specific radioactivity of the putrescine may be adjusted within the range 0.1 - 1.0 µCi / µmol depending on the sensitivity of assay required. Higher levels than this lead to problems of removal of free amine since the filter discs tend to disintegrate with multiple washes and the volumes of TCA required become difficult to handle.

2) Incubate the reaction mixture at 25°C for up to 48 hr, in a stoppered tube.

3) At convenient time intervals remove up to 100 µl of incubation mixture and evenly dispense onto 3 MM (2.1 cm diameter) Whatmann filter discs. Leave at room temperature for 1 min. to be fully absorbed.

4) Immerse the disc(s) in a 10 % w/v solution of ice cold TCA, allow at least 5 ml of TCA soln/disc. Incubate at 4°C for 1 hr.

5) Decant TCA and replace with a second equivalent volume of 5 % w/v TCA and incubate at room temperature for 15 min, with occasional stirring.

6) Repeat previous step, but incubate at 90°C for 35 min.

7) Repeat step number 5)

8) Decant TCA, add equivalent volume of absolute alcohol. Swirl gently for two minutes, and repeat.

9) Decant the ethanol and replace with an equivalent volume of diethylether. Swirl gently for two minutes, decant, and repeat.

10) Allow discs to dry and place into scintillation vials (original sample-loaded side uppermost) and cover with 5 ml of an appropriate scintillation fluid ("OPTIPHASE "MP" was found to be convenient).

11) Count samples in an automated scintillation counter[‡], together with a number of blank discs, and samples of appropriately diluted stock [^{14}C] putrescine.

[‡] Liquid scintillation counting was carried out using the Packard Tri-Carb 2000 CA liquid scintillation analyser (Canberra Packard, Pangbourne, Berks. U.K.). All samples for counting were prepared in fine glass screw topped scintillation vials.

Some proteins (eg catalase) were observed to bind certain amines (ie dansylcadaverine) non-covalently. The conditions used for the determination of covalent binding of amines as part of the filter disc method are so extreme that non-covalently bound amine is unlikely to stay bound. Stoichiometry determinations, following simple desalting of protein by gel filtration [see section 2.2.2.3.(b)], are more prone to error since the native conformation of the amine binding sites are not denatured during the desalting process.

2.2.3.3. DETERMINATION OF THE NUMBER OF GLUTAMINYL RESIDUES WITHIN PROTEINS REACTIVE TOWARDS TRANSGLUTAMINASE

A knowledge of the starting concentration of bovine ß-casein [or other proteins], allows the calculation of a stoichiometry of covalent attachment of putrescine into the protein on the basis of µmol amine/µmol protein [A_{280}: 1mg/ml/cm solution of bovine ß-casein = 0.46 (Thompson and Pepper, 1964)]. Hence if the amount of transglutaminase added to the incubation mixture is known then the specific activity of the enzyme may be calculated.

The determination of labelling stoichiometries on the basis of fluorescence is not recommended since the fluorescence of dansylcadaverine substantially decays over short periods of incubation at 25°C in typical labelling buffers. In the case of proteins modified with [^{14}C] putrescine, samples were diluted and the associated radioactivity measured using a Packard Tricarb 2000 CA liquid scintillation counter. The total amount of putrescine bound to protein was calculated by comparison with the amount of radioactivity associated with a known amount of [^{14}C] putrescine of known specific radioactivity (see Fig 2.5). Stoichiometry calculations derived from the filter disc method (see section 2.2.3.2.) were shown to be > 90 % of that determined by repeated dialysis in the case of ß-casein.

2.2.3.4. A FLUOROMETRIC ASSAY OF TRANSGLUTAMINASE ACTIVITY

This assay is based on the method of Cooke and Holbrook, (1974). Dansylcadaverine (see Fig 2.6.) replaces [^{14}C] putrescine in the reaction mixtures described in section 2.2.3.2.. A stock solution of 1-5 mM was diluted into reaction mixtures to give a final concentration of 0.5 to 1.0 mM amine, as required. Fluorometric assay was not found to be as convenient as the colorimetric (section 2.2.3.1.) or radiolabelling assay (section 2.2.3.2.) for the routine assay of transglutaminase activity. However, the chemical and spectroscopic properties of dansylcadaverine make it particually useful for :

1) separation of modified derivatives of peptides from mixtures (see section 2.2.5.5.).

2) Identification of specific reactive glutaminyl residues within protein substrates of transglutaminase [see section 2.2.6.2.].

The stoichiometry of incorporation of dansylcadaverine into proteins was determined as follows;

PROCEDURE

After the termination of transglutaminase catalysed protein-labelling experiments, the reaction mixtures contain high concentrations of primary amines, EDTA-Ca^{2+} complex, EDTA, DTT. These must be removed from the protein of interest prior to the further analysis of modified material. This was achieved by either:

1) dialysis [see section 2.2.2.3.(a).
2) size exclusion chromatography [see section 2.2.2.3.(b)].

Samples were then analysed using the technique described in section 2.2.2.4. (a) and the amount of dansylcadaverine bound to the protein was determined relative to a standard curve constructed using known amounts of the amine. The formula used for the number of reactive glutaminyl residues within a protein substrate of transglutaminase (S) was calculated using the equation shown in Fig 2.5.

$$\text{stoichiometry [S]} = \frac{\mu\text{mol primary amine } \dagger}{\mu\text{mol protein monomer}}$$

Fig 2.5. THE CALCULATION OF NUMBER OF REACTIVE GLUTAMINYL RESIDUES IN PROTEIN SUBSTRATES OF TRANSGLUTAMINASE [† = maximum observed value]

2.2.3.5. A COMPARISON OF THE COLORIMETRIC, RADIOMETRIC, FLUOROMETRIC ASSAYS OF TRANSGLUTAMINASE

The three types of assay have advantages and disadvantages relative to each other. The colorimetric assay (2.2.3.1.) may be performed quickly (1 hr) and has relatively few steps, compared to the radiolabelling assay which is time consuming and rather expensive. However CBZ-Gln-Gly is only sparingly soluble at high concentrations compared to β-casein at pH values above 7.6.

The colorimetric assay was used for monitoring the purification of transglutaminase and the stability of the enzyme after storage (prior to use in the radiolabelling assays). The radiolabelling assay (see section 2.2.3.2.), and the fluorometric assay (see section 2.2.3.4.) are particularly sensitive and were used in parallel with each other in experiments designed to investigate the specificity of transglutaminase towards novel protein and peptide substrates (see chapter 4). In such experiments the

radioactive and fluorescent nature of the attached amine probe could be used to identify the glutaminyl residue(s) modified by transglutaminase (see section 2.2.6.2.).

2.2.4. THE ASSESSMENT OF STRUCTURAL INTEGRITY OF PROTEIN SUBSTRATES OF TRANSGLUTAMINASE

2.2.4.1. INTRODUCTION

The determinants of transglutaminase specificity lie not only within the primary sequence of proteins, but are also determined by the degree of conformational flexibility in the areas surrounding modification sites (see section 1.7.). Therefore it is important to ensure that substrate sites are correctly folded during their modification by transglutaminase. A number of methods were available for the assessment of different aspects of protein structure, and these are discussed in sections 2.2.4.2. to 2.2.4.6..

2.2.4.2. BY ASSESSMENT OF SPECIFIC ACTIVITY.

Since the enzymic activity of proteins is dependent on the maintenance of their correctly folded tertiary structure, an assessment of the specific activity of an enzyme may be used to judge the integrity of its tertiary structure. This method is most useful when used in conjunction with other techniques such as circular dichroism (see section 2.2.4.3.) and SDS PAGE (see section 2.2.4.4.).

The conditions used for the assay of enzymes tested as transglutaminase substrates in this study are given elsewhere (see appendix II). The

concentrations of stock solutions of proteins were calculated from the published molar absorbance coefficients. Specific activities were calculated in Units/mg/min.

2.2.4.3. CIRCULAR DICHROISM

Circular dichroism (c.d.) spectra of proteins were recorded at 20°C using a JASCO J-600 spectropolarimeter, with assistance from Ms S. Kelly. The cell pathlengths for near u.v. and far u.v. spectra were 1 cm and 0.02 cm respectively, and protein concentrations were typically in the range 0.5 - 1.0 mg/ml. A detailed structural analysis of the proteins tested required measurements in the far u.v. down to 190 nm and for this purpose the spectra were recorded in low strength buffers, typically Tris-HCl or Triethanolamine-NaOH in the range (10 to 25 mM) ie. a little lower in concentration than in the labelling mixtures. Dithiothreitol (DTT) could be added up to a concentration of 10 mM without significant interference with the signal as could 5 mM putrescine and 5 mM Ca^{2+}. Circular dichroism spectra were normalised to zero ellipticity at 250 nm. and mean residual ellipticities calculated at 1 nm intervals using the formula shown in Fig 2.6.

$$\text{MRE } [\Theta]_{m.r.w.} = \frac{\text{MRW x recorded ellipticity}}{\text{[protein] x cell path length x 10}}$$

Fig. 2.6. CALCULATION OF THE MEAN RESIDUE ELLIPTICITY OF A PROTEIN

In Fig 2.6.MRE is the mean residue ellipticity and MRW is the mean residue weight of an amino acid. Recorded ellipticity is measured in degrees. Protein concentration is in g/ml. Cell path length is in cm. The units of $[\Theta]_{m.r.w.}$ are thus;deg. cm^2. $dmol^{-1}$. The secondary structural content was derived using the CONTIN procedure (Provencher and Glöckner, 1981) [see section 2.2.8.3.].

2.2.4.4. SDS-PAGE

Polyacrylamide gel electrophoresis in the presence of S.D.S. was performed by the method of Laemmli, (1970). Polyacrylamide gels were routinely prepared in advance and stored at 4°C, for a maximum of 5 days until required.

For sample preparation protein solutions were mixed with an equal volume of double strength sample buffer containing 10 % v/v 2-mercaptoethanol and boiled for 3 min. After cooling a 1 vol/ 1 vol mixture of 2-mercaptoethanol and 1 % w/v bromophenol blue was added to a final concentration of 5 % v/v and the samples loaded onto gels using a Hamilton syringe. Sample volumes after preparation were in the range of

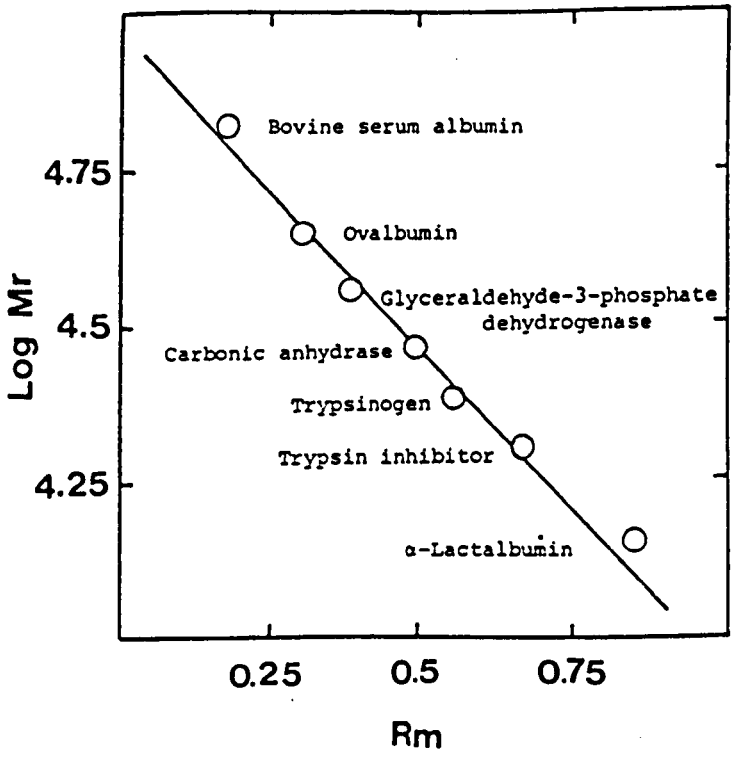


Fig 2.7. MIGRATION OF STANDARD PROTEINS RELATIVE TO THEIR LOG MOL WEIGHT FOLLOWING SDS-PAGE ON A 12 % POLYACRYLAMIDE SEPARATING GEL.

Rm values express the migration of proteins relative to the bromophenyl blue present in each sample. The protein standards (14 kDa to 70 kDa) were purchased from Sigma and prepared as described in section 2.2.4.4., prior to electrophoresis.

10-200 µl and contained 1-20 µg protein. Gels were run at a constant current of 12 mA for 6hr, or 4 mA overnight.

Gels were stained for protein in Coomassie Brilliant Blue R-250 solution and destained in 10 % v/v acetic acid. Protein Mr values were determined from the relative migration of the protein under test relative to a number of protein standards contained in the Dalton mark VII marker set. This marker set contains the following proteins: Bovine serum albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle (35 kDa), Bovine carbonic anhydrase (29kDa) trypsinogen (24kDa), soybean trypsin inhibitor (20 kDa) and α-lactalbumin (14 kDa). A typical calibration curve for these proteins on a 12 % polyacrylamide separating gel is shown in Fig 2.7. Under the conditions used (ie. 5 mM amine) cross-linking of proteins by transglutaminase was not observed to be a predominant reaction, nor were proteinases present in the preparations of proteins tested as substrates.

A variation on this type of electrophoresis was used to produce gels capable of separating peptides. This method required the inclusion of high concentrations of urea, (Anderson et al., 1983). These gels could be electroblotted onto an Immobilon sheet, using a modification of the method of Lin and Kasamatsu, (1983). The dansylcadaverine labelled peptide bands were identified by their fluorescence. These were cut out and retained for microsequencing by the method of Matsudaira, (1987). However HPLC separations of labelled peptides was found to be the most effective way of determining transglutaminase labelling sites due to the greater preparative capacity of columns to accumulate material for analysis (see section 2.2.5.5.).

2.2.4.5. CAPILLARY ZONE ELECTROPHORESIS

This technique separates peptides on the basis of their charge and their size as in gel electrophoresis, but the sample is easily recoverable after analysis. Hence this technique may be used to check the purity of peptides which have been separated from mixtures using HPLC. Samples were run on an ABI 270 A analytical capillary electrophoresis system; capillary diameter 50μm. The buffer used was sodium citrate pH 2.5, (20 mM) (ABI code 400906). The samples were run at 25 kV at 30°C, and peptides were detected by their absorbance at 220 nm. Full scale deflection was 0.01 absorbance units (10 mV equiv). Loading was 5 sec vacuum loading. Traces were recorded using a chart speed of 5 mm/min. This was done by Dr B. Smith of Celltech Ltd.

2.2.5. PREPARATION OF MODIFIED PEPTIDES FOR SEQUENCING

2.2.5.1. INTRODUCTION

Once the number of modification sites within a protein have been estimated the next step is to identify the location of reactive glutaminyl residues within the primary sequence of the substrate. Unless the modified sites are located within the first 20 - 30 N-terminal residues of a protein or peptide which has a free N-terminus it is necessary to cleave the substrate into smaller peptides which have appropriate properties for direct sequencing. To do this the protein is typically :

1) denatured,
2) chemically modified
3) digested using small quantities of appropriate proteinases.

The preparation and purification of peptides which carry the modified residues of their parent protein are described in sections 2.2.5.2. to 2.2.5.5..

2.2.5.2. CARBOXYMETHYLATION OF PROTEIN SUBSTRATES

The procedure used was that of Allen, (1981). Desalted protein was denatured by incubation with a high concentration (6M) of guanidine hydrochloride for 1 hr at 25°C. Samples were then treated with 2 mM DTT, [which will reduce disulphide bonds] and incubated for a further hour. The addition of a 3 mM excess of freshly prepared iodoacetate solution was then added to form the acetylated derivative of the cysteine side-chains. After incubation for a further hour in the dark, the excess iodoacetate was reacted away with a further excess (10 mM) of 2-mercaptoethanol. The chemically modified protein was recovered from the mixture by a second cycle of dialysis against 1 % (w/v) ammonium bicarbonate, pH 8.0. The chemical modification of free cysteine side-chains improves :

1) the solubility of protein (and after proteolysis the solubility of peptide derivatives).

2) overcomes problems associated with the oxidation of cysteine side chains in solution to their sulphonic acid derivatives.

3) prevents the formation of disulphide bridges between material during peptide purification.

2.2.5.3. ENZYMIC DIGESTION OF PROTEIN SUBSTRATES

Tosylphenylchloromethane (TPCK) treated bovine pancreatic trypsin (Sigma) was found to be effective for the generation of modified peptides in high yield in a number of cases at a 1:40 (w/w) ratio of trypsin/protein substrate. Incubations were conducted at 37°C generally for no longer than 4 hr. Other enzymes used included Staphylococcus V8 proteinase and tosyllysylchloromethane (TLCK) treated bovine chymotrypsin.

2.2.5.4. ANALYTICAL HPLC OF THE PRODUCTS OF PROTEOLYTIC DIGESTION

HPLC was used for mapping of peptides produced by the proteolytic digestion of protein substrates of transglutaminase. Peptide products of digestion were separated by application of gradients of increasing solvent concentration (see section 2.2.5.5.), and eluted material was detected from the characteristic absorbance at 218 nm of peptide bonds.

Peptides which were radioactively or fluorescently modified by transglutaminase catalysed reactions were identified by the analysis of a portion of the 1.0 - 3.0 ml of the column fractions which were collected. The sensitivity of detection of dansylcadaverine-modified peptides was increased when necessary by increasing the aperture of the fluorimeter slits.

By careful attention to flow rates (typically 0.2 to 2.0 ml/min) and fraction size (typically 0.2 to 2.0 ml), the modified peptides were identified as specific peaks in the recorded traces. This "fingerprinting" technique was used to determine the optimum digestion conditions for the production of suitable preparations of modified peptides for sequencing. The procedure was then scaled up to a preparative ie. milligram level (see section 2.2.5.5.).

Prior to injection onto HPLC columns samples were firstly analyzed by SDS-PAGE (see section 2.2.4.4.) to avoid the injection of large polypeptides onto the column. Undigested material often has a tendency to precipitate in the solvent systems used, causing column damage. A number of blank gradients were run, prior to analytical separations to ensure that no extraneous material had adhered to the column from previous separations.

2.2.5.5. PREPARATIVE HPLC OF TRANSAMIDATED PEPTIDES

The peptide fragments were separated by reverse phase HPLC on a Vydac C-18 (end capped) column of dimensions 25 cm x 4.6 mm and particle size 5 µm. Peptides were typically eluted from the column by application of a linear gradient between 90 % solvent A, and 70 % solvent B over a period of 30 - 90 minutes. Solvent A consisted of 0.4 % (w/v) aqueous solution of triethylamine, adjusted to pH 2.5 with orthophosphoric acid. Solvent B was solvent A containing 60 % (v/v) propan-2-ol. Modified peptides were located as described in section 2.2.5.4.. Peak fractions were rechromatographed on similar gradients to the initial separation, but using 0.1 % TFA as the modifying agent and acetonitrile as the organic phase (final pH retained at 2.5), for the purposes of peptide sequencing. By rechromatography of purified peptides on relatively shallow gradients of solvent B it was possible to assess the purity of peptides through the sharpness and symmetry of the peak. The labelled peptides were retained for further purification and sequence analysis. Any variation in purification technique is described together with the results presented.

2.2.6. ANALYSIS OF TRANSAMIDATED PEPTIDES

2.2.6.1. INTRODUCTION

Transamidated peptides were characterized by direct peptide sequencing, amino acid analysis and mass spectrometry. These techniques are described in sections 2.2.6.2. to 2.2.6.4..

2.2.6.2. PEPTIDE SEQUENCING

The sequence of the modified peptides were determined using an Applied Biosystems Model 470 gas phase sequencer, with on-line 120A PTH analyzer. Phenylthiohydantoin (PTH) derivatives produced at each cycle were identified by HPLC by reference to standard PTH derivatives of amino acids. Samples of purified peptides were reduced in volume to 50 µl by rotary evaporation using a Gyrovap (care was taken not to reduce samples to dryness) and loaded onto a Biobrene 2 x F2C treated filter. The sample tube was rinsed with 15 µl of glacial acetic acid and this was loaded onto the sequencer. Peptide sequencing of samples derived from transaminated ß-lactoglobulin, phosphoglycerate kinase $His_{388} \rightarrow Gln_{388}$ mutant, and bovine catalase and porcine pepsin was by Edman degredation and was done by Dr. B. Smith of Celltech Ltd., Slough, Berkshire. The analysis of peptides derived from bovine ß-casein was performed by Dr L. Fothergill-Gilmore of the Department of Biochemistry, University of Edinburgh, Scotland.

a). PEPTIDE SEQUENCING OF DANSYLATED PEPTIDES

During the sequencing of dansylated peptides a characteristic gap is observed at that cycle of the sequence determination which releases the

PTH derivative of N-(γglutamyl)dansylcadaverine from the sample. The fluorescence associated with this amino-acid derivative may be detected in the column eluate. A proportion of this material (10 %) was automatically retained after each cycle and this was used for analysis. The 60 µl retained from each sequencer cycle was dried down under vacuum and taken up in a small volume of methanol and spotted onto polyamide sheets and illuminated under u.v. light. The location of dansylated amino acid derivatives within a sequence was confirmed by the characteristic pale yellow glow associated with the dansyl moiety.

b). PEPTIDE SEQUENCING OF [^{14}C]-PUTRESCINE LABELLED PEPTIDES

The [^{14}C]-putrescine derivative of a modified glutaminyl residue was observed to increase retention time of the PTH derivative to approximately that of PTH itself. The modification could be confirmed by scintillation counting of the 60 µl of the column eluate retained after each sequenator cycle. The majority of the radioactivity associated with the original peptide appears at that sequenator cycle which releases the PTH modified N-(γ-glutamyl)putrescine residue. Since the hydrophobic nature of dansylcadaverine invariably caused a large shift in the retention times of modified peptides this amine was generally used in preference to putrescine as a probe of reactive glutamine residues within substrates of transglutaminase.

2.2.6.3. FAB-MASS SPECTROMETRY OF TRANSAMIDATED PEPTIDES

In some cases it was possible to check the calculated Mr of the modified peptide sequence by FAB-mass spectrometry. This technique also served

to show the stable nature of the transglutaminase-mediated isopeptide bond formed between the protein of interest and the dansylcadaverine. Peptide mass determination was carried out on a VG Trio 3 Triple Quad mass spectrometer with an Ion Tech Xenon Fast Atom Bombardment gun. The scan mass range was from Mr 1500 - 2500, calibrated with Caesium Iodide clusters. Samples of 500 ng modified peptide were dissolved in 5 µl of 10 % (v/v) aqueous acetic acid and 1 µl added to thioglycerol on the target. This work was done by Mr T Gibson, Xenova Ltd, Slough, Berkshire.

2.2.6.4. AMINO ACID ANALYSIS OF LABELLED PEPTIDES

Peptide and protein samples were subjected to prolonged acid hydrolysis and the products analyzed using an automated (Applied Biosystems) amino acid analyzer. This work was done by the technical staff at Celltech Ltd. Slough, Berks.

2.2.7. THE DETERMINATION OF SUBSTRATE SITES FOR TRANSGLUTAMINASE WITHIN PROTEIN SUBSTRATES

2.2.7.1. INTRODUCTION

The conditions used to test the substrate properties of a protein or peptide are similar to those used for the radiolabelling assay (see section 2.2.3.2.) and the fluorescence assay (see section 2.2.3.4.) of transglutaminase. The substrate to be tested was substituted for ß-casein in the incubation mixture at a concentrations which were typically in the range of 3.0 - 6.0 mg/ml. The reducing agent DTT was also added to incubation mixtures typically in the range of

1.0 - 5.0 mM, as required, see section 3.4.2.. Whenever possible the components of the incubation mixture were dissolved in buffers which:

1) resembled that in which the data for the construction of a folded structural model had been derived,

2) were compatible with established assay conditions, (if the test protein was known to have an enzymic activity dependent on a highly folded structure).

Tris-HCl (50mM) pH 7.0 - 8.0, and triethylamine-NaOH (50 mM) pH 7.0 - 8.0, were found to be satisfactory in most cases.

2.2.7.2. EXPERIMENTAL PROCEDURE

Protein labelling was carried out both with [^{14}C]-putrescine and dansylcadaverine in parallel. Samples were removed at regular intervals for;

1) determination of stoichiometry of modification by the filter disc procedure (see section 2.2.3.2.) and fluorometry (see section 2.2.3.4),
2) Dilution and enzyme assay (see section 2.2.4.2.),
3) electrophoretic analysis on SDS-gels (see section 2.2.4.4.),
4) Dilution (usually by three to six fold) and analysis of secondary structural integrity by circular dichroism (see section 2.2.4.3.).‡
5) Desalting by size exclusion chromatography or by dialysis (see section 2.2.2.3.).

6) for the calculation of labelling stoichiometries (2.2,3.4.).

7) final analysis of modification site(s) by proteolytic digestion, and purification by HPLC of suitable sized peptides for peptide sequencing (see sections 2.2.5.5.).

8) Characterization of labelled peptides by direct sequencing (see section 2.2.6.2.), amino acid analysis (2.2.6.4.) and mass spectrometry (see section 2.2.6.3.).

9) The location of the modified glutaminyl residue was located within the three dimensional structure of the protein by viewing computer generated graphical models of the substrates (see section 2.2.8.1).

† Circular dichroism could not be performed on samples which contained excess dansylcadaverine due to the high absorbtivity of this compound over the 220 - 400 nm range.

The activity of transglutaminase was always tested by the colorimetric method (2.2.3.1.) approximately 1 hr prior to execution of these large scale (and hence expensive) protein labelling experiments. The buffers used for the colorimetric assay of transglutaminase activity were constructed to resemble the buffer which was used for the following protein modification experiment. The substrate properties of known transglutaminase substrates eg. ß-casein were examined in parallel with any new protein under test, together with proteins known to be poor substrates eg. pancreatic ribonuclease A (bovine) [see section 4.3.1.1]. Each labelling experiment was performed in parallel with a control lacking transglutaminase. All proteins for which the modification sites were determined were tested a minimum of three times using the above procedure. Direct amino acid sequence determinations were only performed once on each of the purified peptides

examined. Care was taken to ensure that the maximum observed stoichiometries of incorporation of amine into proteins by transglutaminase correlated satisfactorily with the number of directly determined modification sites.

2.2.8. THE USE OF COMPUTER BASED TECHNIQUES IN THE DETERMINATION OF TRANSGLUTAMINASE SPECIFICITY

2.2.8.1. THE USE OF COMPUTER GRAPHICS

Computer facilities at Aberdeen and Glasgow universities, and Celltech Ltd. Slough, Berks. were used for the viewing of the x-ray structures of a number of proteins from the Brookehaven database. The molecular models were generated using the "QUANTA" programme from Polygen and were displayed using an IRIS 3130 works station from Silicon graphics. The BIPED procedure (Akrigg et al., 1988) was used to assess the relative accessibilities of glutaminyl side chains at the surface of proteins within this database. This data was obtained with the assistance of Dr L. Sawyer. Energy minimization of peptide structures were done using the CHARMm programme. The FRODO programme was used for the superposition of the 3-D structures of a number of protein and polypeptide structures of transglutaminase. The structure of bovine β-lactoglobulin was stereoscopically viewed at Edinburgh University, by kind permission of Dr. L. Sawyer.

2.2.8.2. SECONDARY STRUCTURAL PREDICTION OF PROTEINS AND PEPTIDES USING COMPUTER ALGORITHMS

A range of methodologies is available for the prediction of local secondary structure within peptide sequences. The method of Chou and Fasman, (1974 a,b, 1977) was used to predict the secondary structure of transglutaminase. Many of the other structural predictions on shorter sequences were done using the programme PREDICT (Eliopoulos et al., 1982) as modified by E. Eliopoulos (see Sawyer et al., 1986, 1987) to include up to eight different predictions of α-helix, ß-sheet, ß-turn, or other structure. For an α-helix or ß-strand to be predicted, it was required that more than three methods should give the same prediction over four or more contiguous peptide bonds. For ß-turns the requirement for contiguous residues was dropped. All secondary structural predictions were done by Dr L. Sawyer, University of Edinburgh (see section 9.2).

2.2.8.3. THE PREDICTION OF SECONDARY STRUCTURE FROM DATA OBTAINED FROM CIRCULAR DICHROISM

Circular dichroism is widely used for studying the conformation and conformational changes in proteins (see section 2.2.4.3.). The usual measure of secondary structure is achieved by approximation of the mean residue ellipticity at wavelength λ simply by linear superposition of a small set of N_r reference spectra r_i (λ) each of which is supposed to be characteristic of a particular conformational class. The classes α-helix, ß-sheet and remainder (ie. all that does not belong to the other classes) were used.

$$y(\lambda) = \sum_{i=1}^{N_f} f_i r_i(\lambda)$$

Fig 2.8. Equation for the calculation of mean residue ellipticity at a wavelength. Abbreviations: f_i is the fraction of residues in class i, and the $r_i(\lambda)$ values are previously determined from c.d. spectra of model polypeptides (Greenfield et al., 1987) or globular proteins (Saxena and Watlaufer, 1971; Chen et al., 1972) whose secondary structures are known.

The mean residue ellipticity of a protein at a certain wavelength may be calculated from the equation shown in Fig 2.8.. No single reference spectrum can accurately represent all the members of the three conformationally defined classes, (Chang et al., 1978) since attempts to take all these parameters into account would be too complex using the least squares estimates of f values and they would be unstable to experimental error. Provencher and Glöckner, (1981) have developed a simple constrained regularization procedure which yields solutions for f values which are stable to experimental error even when there are many parameters. C.D. spectra are interpreted in terms of a linear combination of the c.d. spectra of sixteen proteins of known secondary structure. The problem of defining reference spectra is thus avoided.

2.2.8.4. THE USE OF COMPUTERS FOR SEARCHING DATABASES OF PROTEIN SEQUENCES

Computer based methods were used to search databases for peptide sequences which resembled the known substrate sites of transglutaminase, and the statistical frequency of occurrence of sequence motifs associated with transglutaminase substrate sites. This work was done using the I.C.L., 64 x 64 Distributed Array Processor (DAP; Flanders *et al.*, 1978) as described by Coulson *et al.*, 1987). The programmes for inexact string-matching were developed by Lyall *et al.*, (1986) : this has implemented the " Best Local Similarity " algorithm of Smith and Waterman (1981). This work was done by Dr L. Sawyer, Department of Biochemistry, and Dr A. Coulson, Department of Molecular Biology, University of Edinburgh.

CHAPTER THREE: PURIFICATION OF TRANSGLUTAMINASE FROM GUINEA PIG LIVER

3.1. INTRODUCTION

Transglutaminase was purified from guinea pig liver using a procedure which combined features of the methods of Connellan et al., (1971), and of Folk and Cole, (1966). The pH optimum for the enzyme, and the effect of reducing agents on transglutaminase activity were investigated and compared with the commercially available (Sigma) enzyme for the substrates CBZ-Gln-Gly and β-casein. The substrate properties of a number of proteins of known sequence (and in some cases tertiary structure) were investigated. Some model systems were identified for extending the investigation of the basis of transglutaminase specificity.

3.2. METHODS

Freshly excised guinea pig livers were stored in 0.25 M sucrose soln for up to one week at 4°C prior to use, otherwise they were stored at -20°C until required. Transglutaminase was purified from this material using a combination of ion exchange, and size exclusion chromatography to such a stage that only a single band was observed on a 10 % polyacrylamide gel following the analysis of samples as described in section 2.2.4.4.

The apparent molecular weight was estimated to be 80 kDa. Protein concentrations were measured using the Coomassie Blue binding method of Sedmak and Grossberg, (1977), see section 2.2.2.2. The purification was carried out at 4°C unless otherwise stated, and involved the following steps:

1) 45 g of fresh livers were cut up into 1.0 cm^3 cubes and suspended in 0.25 M sucrose solution (at 4°C) added to give a final volume of 112 ml. This mixture was homogenized quickly using a teflon in glass homogenizer.

2) The homogenate was centrifuged at 20,000 rpm. for 40 min.

3) 20 ml of sodium acetate (100 mM) was added to 100 ml of supernatant to a final concentration of 16.6 mM.

4) The sample was adjusted to pH 5.0 with 1.0 M acetic acid over two minutes, with stirring.

5) The cloudy mixture from step 4) was centrifuged for fifteen minutes at 20,000 r.p.m. and the supernatant was retained.†

6) 25ml of supernatant was then applied to a DEAE-Sephadex column (dimensions 8.0 x 20 cm) and eluted with a linear gradient of 0.1 - 0.5 M KCl in tris-HCl (5 mM) pH 7.5. containing EDTA (2 mM). Column fractions were assayed initially on the basis of esterase activity of fractions towards p-nitrophenyl acetate (Gross and Folk, 1973) in order to establish the approximate location of transglutaminase (phosphatases, and proteinases can also catalyse this reaction). Transglutaminase activity was then confirmed by the colorimetric hydroxamate procedure (see section 2.2.3.1.) of Folk and Cole, (1966).

7) The enzyme rich fractions were combined and ammonium sulphate was added to give 60 % saturation, with gentle stirring over fifteen minutes.

† This was adjusted to pH 7.5 by the dropwise addition of NaOH (aq) with stirring.

The solution was then centrifuged at 12,000 g for twenty minutes and the pellet collected. The supernatant was adjusted to 80 % saturation and the previous centrifugation step was repeated.

8) The pellets from step 7 were redissolved in 5 ml of Tris-acetate (10 mM) pH 6.0, EDTA, (2 mM), KCl, (0.16 M), and applied to a sephacryl S-300 column † equilibrated with the same buffer. This buffer was also used to elute proteins from the column, 2.6 ml fractions were collected. The fractions were assayed as described in step 6). The peak fractions were then analyzed by SDS-PAGE (see section 2.2.4.4.) to assess the purity of samples. Samples were stored in various ways in order to assess stability.

3.3. RESULTS AND DISCUSSION

3.3.1. PURIFICATION OF TRANSGLUTAMINASE

A purification table for transglutaminase is shown in Table 3.1.. Transglutaminase was estimated to represent more than 95 % of the final material as judged by staining of SDS-PAGE gels with Coomassie brilliant blue G250. The enzyme purified by this method was found to be identical in molecular weight (~80 kDa), and homogeneity to the commercial source. Both the laboratory purified and commercial sources of transglutaminase were judged free from contamination by proteinases when bovine ß-casein was used as a substrate [(Northrop et al., (1948) and Nomoto et al., (1960)]. The variation in yield of enzyme was related to the period and temperature of storage of livers prior to homogenization. The best results were obtained using livers which had not been frozen, and stored in sucrose at 4°C for less than two weeks prior to use.

† Dimensions 100 cm X 3.5 cm

TABLE 3.1.

Purification of transglutaminase

Based on 45g of guinea pig liver

PURIFICATION STEP	PROTEIN (g)	Specific activity (units/mg)	yield (%)
Homogenate	100.0	0.026	100
Supernatant	35.0	0.015	20
pH /centrifugation step. resuspended pellet	10.5	0.030	12
DEAE-sephadex chromatography	1.5	0.10	5.7
Gel filtration	0.03	2,0	2.3

The fractions were assayed on the basis of the CBZ-Gln-Gly colourimetric assay described in section 2.2.3.1..

livers which had not been frozen, and stored in sucrose at 4°C for less than two weeks prior to use.

3.3.2. THE STABILITY AND CONDITIONS OF STORAGE OF TRANSGLUTAMINASE

When transglutaminase was stored in aqueous buffered solution at pH 6.0 at 4°C the specific activity of transglutaminase was found to decrease by approximately 50% in two weeks. Storage at -20°C did not significantly improve the stability of samples. Freeze/thawing reduced the specific activity of samples further. Freeze dried samples had far greater stability. The commercial source of transglutaminase was prepared from a lyophilized powder as a 1 mg/ml solution in double distilled water and freeze dried in 100 µl aliquots in 1 ml Eppendorf tubes. The enzyme was stored at -20°C for at least six months with < 10% loss of activity.

Since the supply of guinea pig livers could not be guaranteed and the commercial source was shown to possess activity directly resembling that in previously published reports the commercial source of enzyme was used for the protein modification experiments described in chapters 4 to eight. Different batches of this enzyme were shown to have consistent properties.

3.4. THE ACTIVITY OF TRANSGLUTAMINASE (CONTROL STUDIES)

3.4.1. INTRODUCTION

The literature shows that transglutaminase activity has been routinely assayed in mixtures which contain reducing agents such as DTT (see section 1.6.6.). Since reducing agents may cause conformational change

in disulphide bonded proteins, it was decided to investigate the possibility that the enzyme could be used in the absence of such compounds. The effect of inclusion of DTT in standard assays of transglutaminase activity using CBZ-Gln-Gly as acyl donor was investigated. The structure of this substrate is not thought to be affected by the inclusion of reducing agents.

3.4.2. THE EFFECT OF REDUCING AGENTS ON TRANSGLUTAMINASE ACTIVITY

3.4.2.1. METHODS

Standard reaction mixtures (as used by Folk and Cole, 1966) were set up for the assay of transglutaminase activity (see section 2.2.2.1.). Dithiothreitol (DTT) was freshly prepared as a 100 mM stock solution in buffer (Tris-HCl, 50 mM, pH 7.5) and diluted into the mixtures to a final concentration of 0.0, 1.0, and 5.0 mM respectively. A stock solution (1 mg/ml) of guinea pig liver transglutaminase was freshly prepared [by addition of double distilled water to the lyophilized enzyme], and 20 µl of this solution was added to each mixture. Control mixtures from which either transglutaminase or Ca^{2+} were omitted were prepared in parallel. The experiment was conducted in duplicate and repeated three times.

3.4.2.2. RESULTS AND DISCUSSION

Approximately 30 % of the activity of the enzyme was lost in the absence of DTT (see Fig 3.1.). Folk and Cole, (1965) have reported a 42 % reduction in the activity of transglutaminase over a 1.0 hr period in the absence

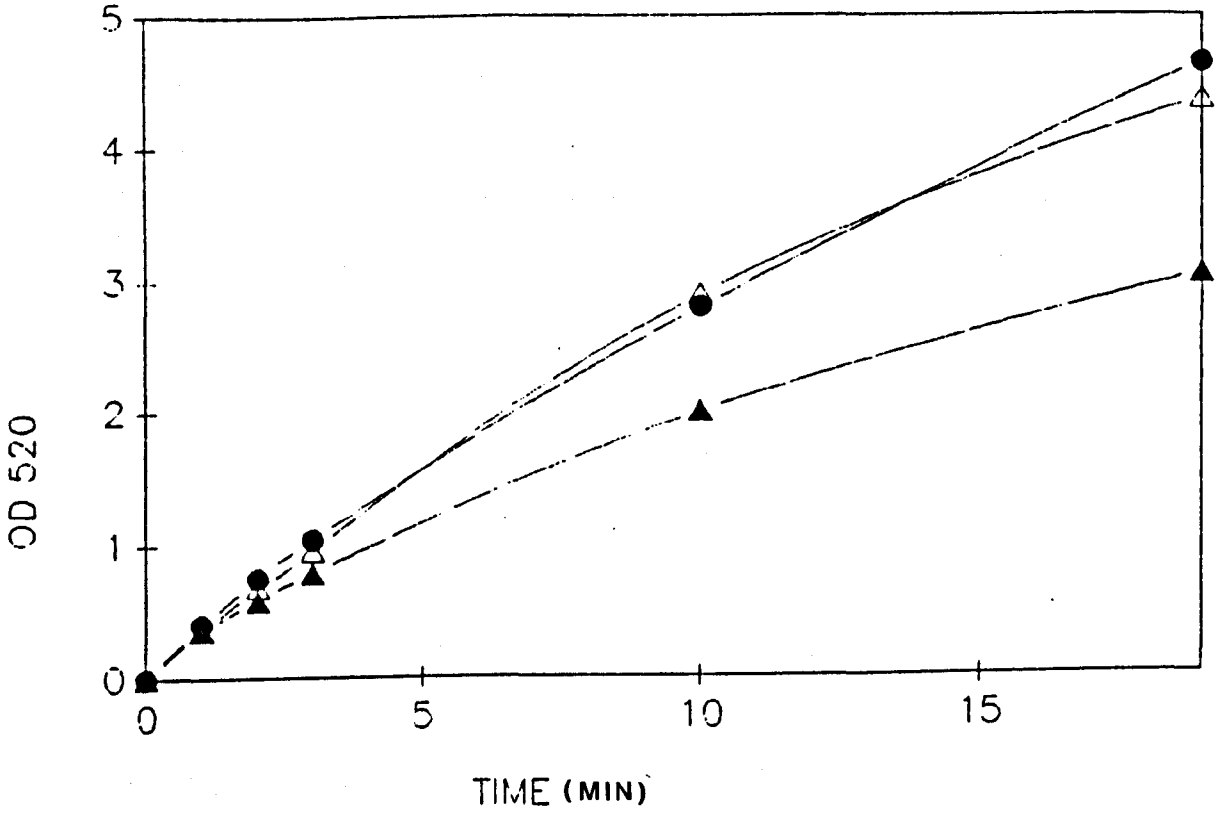


Fig 3.1. THE EFFECT OF DTT ON THE ENZYMIC ACTIVITY OF TRANSGLUTAMINASE.

Details concerning the general conditions used for the assay of transglutaminase towards CBZ-Gln-Gly are given in the text. Where DTT was omitted water was supplemented so that all the reactions occurred in the same volume. The activity of transglutaminase was monitored in the absence (▲), and in the presence of DTT at concentrations of 1.0 (●) and 5.0 mM (△). Samples were incubated at 37°C. Transglutaminase was totally inactive in the absence of added calcium ions. Tubes were prepared in duplicate and the experiment was repeated three times.

of 5.0 mM glutathione. Full activity was observed with 1.0 mM DTT. Raising the concentration to 5.0 mM had no further effect. From the results of chemical modification experiments (Folk and Cole, 1966; Boothe and Folk, 1969) transglutaminase does not appear to contain any disulphide bonds in its native state, although many cysteine residues are titratable, and hence located at the surface of transglutaminase. The reason for the effect of DTT could be due to:

1) protection of the active site Cys from oxidation,
2) the prevention of disulphide interchange within transglutaminase.

3.4.3. pH ACTIVITY PROFILE OF TRANSGLUTAMINASE

3.4.3.1. METHODS

Assay mixtures were prepared as described in section 2.2.3.1., with the exception that triethanolamine HCl (100 mM) [titrated to the appropriate pH with 1 M NaOH (aq)] was used to buffer the mixtures over a pH range from 5.5 - 9.0. Samples were incubated for 1 hr in the presence and absence of DTT (1 mM) and assayed by the method of Folk and Cole (1966). The conformation and charged state of CBZ-Gln-Gly is not expected to be strongly affected by the pH of the assay mixture and hence is considered a suitable substrate for investigating the pH activity profile of transglutaminase.

3.4.3.2. RESULTS AND DISCUSSION

The specific activity of transglutaminase was calculated to be 0.65 U mg^{-1} min^{-1} at pH 6.0. The effect of increasing pH on the activity of transglutaminase is shown in Fig 3.2. The highest initial rate of enzyme activity was at pH 8.0 using CBZ-Gln-Gly as substrate. This compares well with the pH optimum observed by Clarke et al., (1959) and Folk and Cole, (1966). The pH optimum (6.5) reported by Borsook et al.,(1953) may have been the result of the use of phosphate buffers in the Ca^{2+}-dependent lysine incorporating system.

CBZ-Gln-Gly was not fully soluble above pH 8.0 since a slight cloudiness was observed in the mixture. The true pH optimum is likely to be closer to 8.5, which was that observed when β-lactoglobulin and β-casein were tested as substrates (see chapters 5 and 7). This is comparable with the optimum pH of 8.5 observed for the incorporation of dansylcadaverine into heat treated guinea pig liver proteins observed by Clarke et al.,(1959).

3.5. CONCLUSIONS

The laboratory purified transglutaminase appeared to be identical to both transglutaminase purified by Connellan et al., (1971) and to the commercially available source. Hence it appears reasonable to compare the results obtained as part of this thesis with the published literature. The observation that DTT is not essential for activity, allows transglutaminase to be tested on protein substrates in the absence of reducing agents. A lack of protein modification under such conditions is most likely to be due to the poor substrate properties of the protein under test, rather than inactivation of transglutaminase.

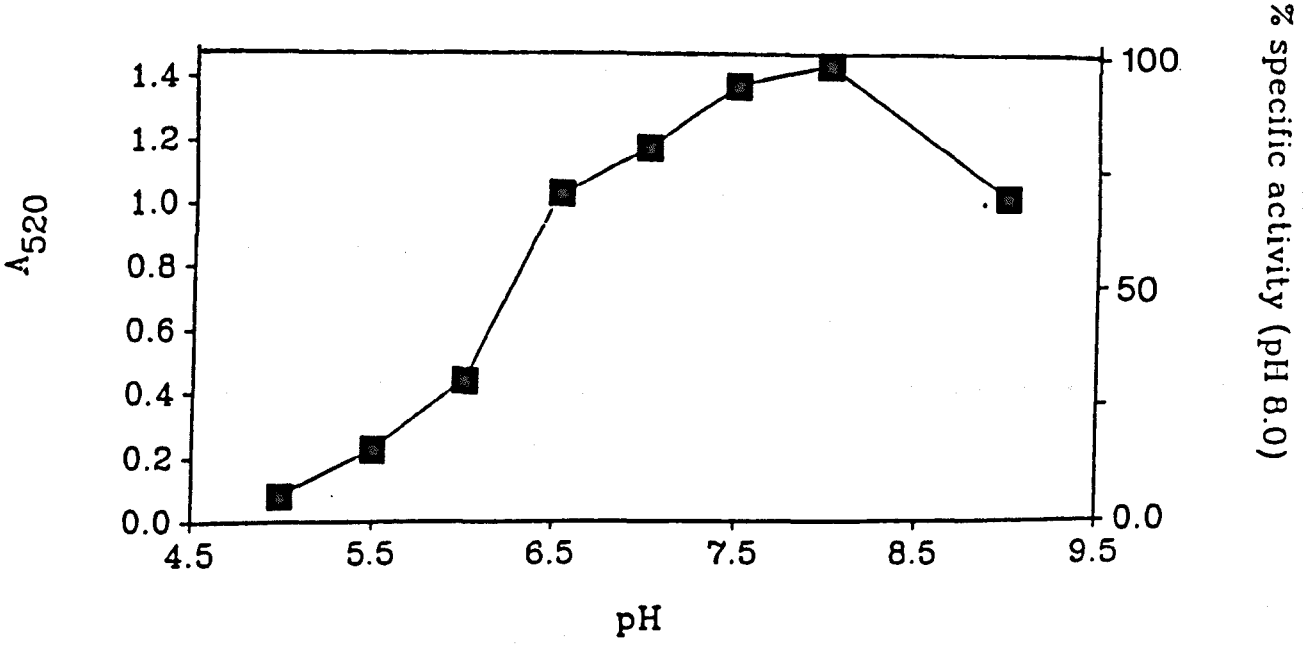


Fig 3.2. THE EFFECT OF pH ON THE ACTIVITY OF TRANSGLUTAMINASE

The effect of pH on the enzymic activity of transglutaminase using CBZ-Gln-Gly and hydroxylamine as substrates was tested using the general reaction and buffer system described in section 2.2.3.1.. Tubes were prepared in duplicate and the experiment was repeated twice.

CHAPTER FOUR PRELIMINARY SCREENING OF PROTEINS AS SUBSTRATES

4.1. INTRODUCTION

A review of the literature (see section 1.7.) concerning the substrate properties of a number of peptides and proteins for transglutaminase has suggested some recurrent themes in the sequences which surround reactive and non-reactive glutaminyl residues. Few globular protein substrates of transglutaminase have been structurally characterized, and investigation of the substrate properties of these compounds may provide greater insights into the relationship which exists between:

1) a primary amino acid sequence,
2) its folded structure,
3) its conformational flexibility,
4) local steric constraints about glutaminyl residues,
5) the substrate properties of glutaminyl residues within the sequence.

A number of proteins were screened for substrate properties towards transglutaminase and the degree of modification of these proteins was calculated as described in section 2.2.2.3.. Protein concentrations were calculated on the basis of the known absorbance coefficients for 1 mg/ml solutions of these proteins as recommended by the suppliers of these materials (see section 2.1.1.). This chapter identifies some model systems for understanding the basis of the specificity of transglutaminase.

4.2. THE IDENTIFICATION OF SUBSTRATES OF TRANSGLUTAMINASE

4.2.1. METHODS

A number of proteins and peptides were tested as substrates for transglutaminase by their substitution for bovine ß-casein in standard transglutaminase assay mixtures described in section 2.2.3.2.. In many cases substrate properties were tested in the presence and absence of reducing agents (see section 3.4.2.).

4.2.2. RESULTS AND DISCUSSION

Table 4.1. shows those proteins which did not appear to have substrate properties. The results of these studies illustrate the general rule that denatured proteins tend to have greater numbers of substrate sites than their folded counterparts, eg. antibody IgG B.72.3 (raised against mucein) lacked substrate properties, unless partially denatured by DTT treatment. Hence the most compact proteins (as judged by their X-ray structures) tended to be the least effective as substrates eg. ovalbumin (Stein *et al.*, 1990), ribonuclease A (Wlodawer and Sjolin, 1983), and cytochrome c (Bushnell *et al.*, 1990).

Table 4.2. shows the proteins which appeared to have substrate properties, together with the observed stoichiometry of modification, and references relating to the structure of the compound. Several globular proteins are evident in this list, eg. bovine ß-lactoglobulin, a mutant of yeast phosphoglycerate kinase (PGK) [where His_{388} was mutated to Gln_{388}], rabbit muscle aldolase, yeast alcohol dehydrogenase, lactate dehydrogenase and bovine liver catalase.

Table 4.1.

Proteins tested which have poor substrate properties for transglutaminase

PROTEIN	MAXIMUM INCORPORATION OF PUTRESCINE AT pH 7.6 (mol amine/mol protein) +DTT (1 mM)	-DTT	COMMENTS
Ribonuclease A Type IIa (bovine pancreas)	0.05	0.05	See section 4.3.1.
Ribonuclease A Type XIIa (bovine pancreas)	0.50	0.20	See section 4.3.1.
Cytochrome c (horse heart)	0.05	0.05	See section 4.3.3.
Ovalbumen (Chicken)	0.10	0.05	Compact structure
Lysosyme (Chicken egg white)	0.05	0.05	See section 4.3.4.
DHQase (E. Coli)	0.05	0.05	Substrate covalently bound
	0.20	0.15	No bound substrate
BSA (serum albumen)	0.40	0.20	Fairly compact tri-domain structure
PGM (yeast)	0.20	0.15	Globular protein with Gln-containing C-terminal tail.
PGK (phosphoglycerate kinase)	0.15	0.10	Globular protein with many exposed Gln residues
Haemoglobin	0.25	0.10	Globular protein with Gln no 2 exposed

The labelling conditions were as described in section 2.2.3.2. ie. pH 7.5 and incubations were performed at 25°C for approximately 20 h in all cases. At higher pH values more Gln residues became reactive in some cases eg. in the presence of 5 mM DTT up to 6.0 mol putrescine/mol protein could be incorporated into porcine pepsin.

Table 4.2.

Proteins determined to have substrate properties for transglutaminase

PROTEIN	MAXIMUM INCORPORATION OF PUTRESCINE AT pH 7.6 (mol amine/mol protein) (+DTT)	(-DTT)	COMMENTS
β-lactoglobulin (bovine)	2.0	0.15	See chapter 5
(His388-Gln388) mutant of PGK	1.2	1.0	See chapter 6
bovine β-casein	2.0	1.5	See chapter 7
porcine pepsin	3.5	1.0	See chapter 8
bovine liver catalase	3.0	2.0	See chapter 8
fibrin peptide	2.2	1.9	See chapter 8
ADH (yeast)	2.0	1.3	no effect on ADH activity
LDH (rabbit muscle)	2.0	1.3	no effect on LDH activity
aldolase (rabbit muscle)	2.0	1.5	effect not tested
IgG B72.3	2.0	0.3	effect on antigen binding not tested (vs mucein)

The incubation conditions for these experiments were identical to those described in Table 4.1.

Some of the substrates tested have previously been shown to have substrate properties (Brenner and Wold, 1978). In some cases the maximum observed stoichiometry of labelling was lower than previously reported eg. only 1.0 - 2.0 mol amine/mol bovine ß-casein represented only 30 % of the total incorporation (5 - 8 mol dansylcadaverine/mol succinylated ß-casein) observed by Yan and Wold, (1984). In other cases higher stoichiometries of labelling were observed than has been reported by other workers eg. Brenner and Wold, (1978) observed only 2.0 mol amine/mol pepsin in the presence of DTT compared to > 5.0 mol/mol incorporation reported in this work. These differences probably reflect the shorter period of incubation of samples in the former case.

4.3. STRUCTURAL FEATURES OF GLOBULAR PROTEIN SUBSTRATES

The location of reactive glutaminyl residues may be predicted within some of the protein substrates of transglutaminase listed in Table 4.2, since they are the only ones which appear to be exposed in models constructed on the basis of X-ray crystallographic data. However, the specificity of transglutaminase seems to depend on more than simply the degree of exposure of glutaminyl residues, since in most proteins, more glutaminyl residues are exposed than the maximum stoichiometry of labelling. The substrate properties of some of the proteins tested are discussed on the basis of their structures in sections 4.3.1. to 4.4.3..

4.3.1. GLOBULAR PROTEINS WITH POOR SUBSTRATE PROPERTIES FOR TRANSGLUTAMINASE

4.3.1.1. BOVINE PANCREATIC RIBONUCLEASE A

The lack of substrate properties of the native RNAase A confirms the findings of Yan and Wold, (1984). A consideration of the X-ray structure of bovine pancreatic ribonuclease A (Wlodawer and Sjolin, 1983) shows that Gln_{28} has a relative accessibility greater than 0.67 [using the criteria of Chothia (1975)] and hence is fully extended in the crystal structure (see Fig 4.1.). The lack substrate properties of ribonuclease A is therefore likely to be due to :

1) a lack of sufficient flexibility about exposed glutaminyl residues,

2) other discouraging features, such as the presence of unfavourable amino acid side chains in the linear sequence surrounding these residues.

RNAase is disulphide bonded in the regions adjacent to the exposed glutaminyl residues. The conformational restriction of residues surrounding potential substrate sites is unlikely to allow sufficient local conformational flexibility for an extended region of the region of polypeptide chain surrounding the glutaminyl residues to bind efficiently at the active site of transglutaminase. It may be significant that there is a high concentration of positive charge on the C-terminal side of Gln_{28} due to the presence of Lys_{31} and Arg_{33}. The sequences surrounding the glutaminyl residues of RNAase are shown in Table 4.3. On the basis of amino acid sequence surrounding the glutaminyl residues within ribonuclease, certain

residues are not expected to have high reactivity even if the protein is denatured eg. Gln_{28}, and Gln_{55}. Yan and Wold, (1984) have reported that RNAase may be modified to a stoichiometry of 3 mol amine/mol protein in samples which were firstly oxidized with performic acid and then succinylated. This could be due to the reduction of positive charge surrounding glutaminyl sites rather than simply being due to increased exposure of reactive side chains.

Table 4.3. SEQUENCES FLANKING GLUTAMINYL RESIDUES IN RIBONUCLEASE A

RESIDUE	AMINO ACID SEQUENCE	RESIDUE
5	A-A-K-F-E-R-Q-H-M-D-S-S-T-S	18
22	S-S-N-Y-C-N-Q-M-M-K-S-R-N-L	35
49	E-S-L-A-D-V-Q-A-V-C-S-Q-K-W	62
54	V-Q-A-V-C-S-Q-R-N-V-A-C-K-N	67
63	V-A-C-K-N-G-Q-T-N-C-Y-Q-S-Y	76
69	Q-Q-T-N-C-Y-Q-S-Y-S-T-M-S-I	82
95	C-A-Y-K-T-T-Q-A-N-K-H-I-V-COOH	108

Positively charged residues are in triangles and negatively charged residues are in circles.

A higher stoichiometry of modification was observed when RNAase XIIa (Sigma) was tested as a substrate (see Table 4.1). This protein is especially low in phosphate content and hence not only is the molecule thought to be less rigid but also Gln_{11} is thought to be more exposed (Campbell et al, 1987)..

4.3.1.2. CONCANAVALIN A

A consideration of the X-ray structure of concanavalin A (Reeke et al., 1975) [see Fig 4.2], shows that the single highly exposed glutaminyl residue (Gln_{43}) is located in a ß-bend. The lack of reactivity may be explained by ;

1) a lack of sufficient flexibility in this region. Residues Gln_{43} to Lys_{46} form a ß-bend which may have considerable conformational stability

2) the location of a positively charged residue, Lys_{46} in the C-terminal region (see Fig. 4.3.).

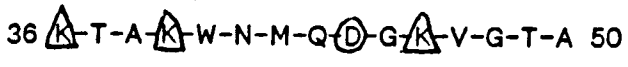
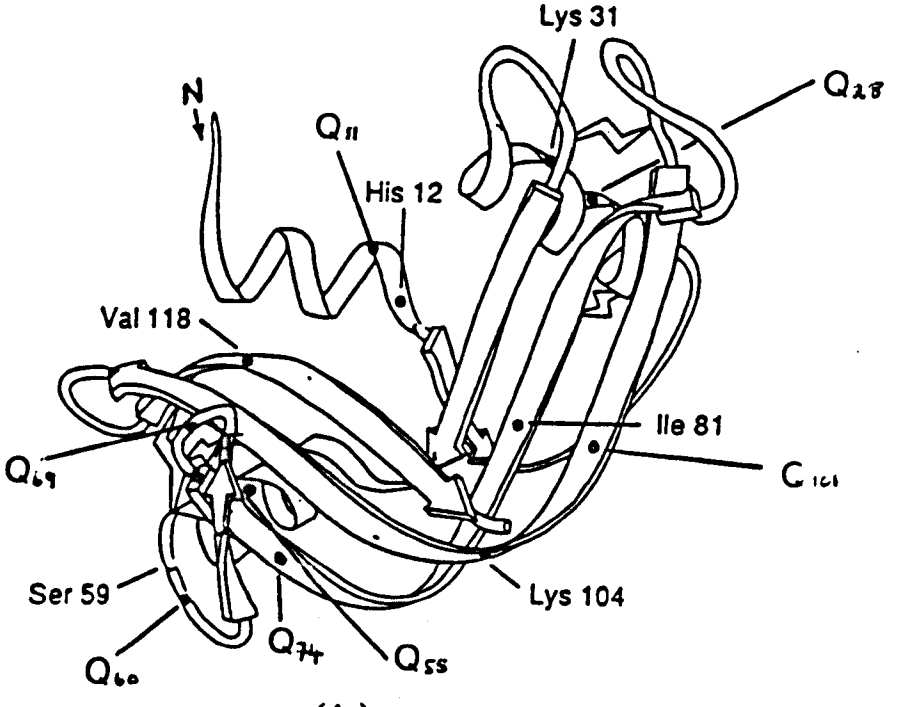


Fig 4.3. SEQUENCE SURROUNDING EXPOSED GLN_{43} IN CONCANAVALIN A

When the published amino acid sequence of con A (Becker et al., 1976) is matched to the crystal structure, glutaminyl residues Gln_{122}, Gln_{132}, Gln_{137}, and Gln_{143} are not sufficiently exposed to act as substrates for transglutaminase as they are either internal or at the dimer interface or the accessibility of these residues being less than 0.67. Alternatively the side-chain of glutamine Gln_{43} is located midway along a stretch

	1				5					10					15		
1	LYS	GLU	THR	ALA	ALA	ALA	LYS	PHE	GLU	ARG	GLN	HIS	MET	ASP	SER	15
16	SER	THR	SER	ALA	ALA	SER	SER	SER	ASN	TYR	CYS	ASN	GLN	MET	MET	30
31	LYS	SER	ARG	ASN	LEU	THR	LYS	ASP	ARG	CYS	LYS	PRO	VAL	ASN	THR	45
46	PHE	VAL	HIS	GLU	SER	LEU	ALA	ASP	VAL	GLN	ALA	VAL	CYS	SER	GLN	60
61	LYS	ASN	VAL	ALA	CYS	LYS	ASN	GLY	GLN	THR	ASN	CYS	TYR	GLN	SER	75
76	TYR	SER	THR	MET	SER	ILE	THR	ASP	CYS	ARG	GLU	THR	GLY	SER	SER	90
91	LYS	TYR	PRO	ASN	CYS	ALA	TYR	LYS	THR	THR	GLN	ALA	ASN	LYS	HIS	105
106	ILE	ILE	VAL	ALA	CYS	GLU	GLY	ASN	PRO	TYR	VAL	PRO	VAL	HIS	PHE	120
121	ASP	ALA	SER	VAL												135

(a)

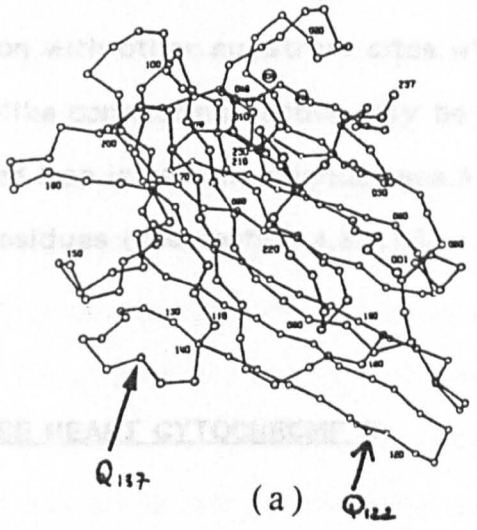


(b)

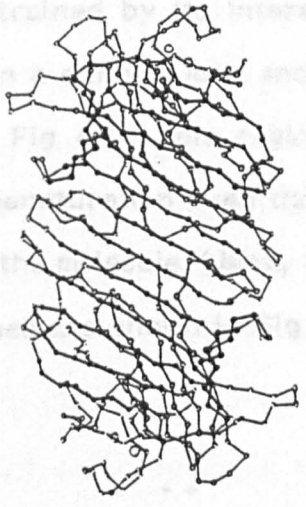
Fig 4.1. THE STRUCTURE OF BOVINE PANCREATIC RIBONUCLEASE A

(a) The amino acid sequence of bovine pancreatic ribonuclease A (Smyth and Moore, 1963).

(b) Schematic drawing of the the trace of the polypeptide in a subunit of bovine pancreatic ribonuclease A. The structure was taken from (Campbell and Petsko, 1987). Regions of β-sheet are arrow-like bars and α-helix are coils. A variety of residues are labelled in order to orientate the reader with respect to the overall folding of the polypeptide chain.



(a)



(b)

Fig 4.2. THE STRUCTURE OF CONCANAVALIN A

(a) Schematic drawing of the backbone α-carbon trace of the polypeptide in a subunit of concanavalin A from jackbean in standard orientation ie. viewed down the x-axis. The metal binding regions is at the top, and the carbohydrate binding region at the lower right. The bottom represents the dimer interface. The approximate position of the exposed glutaminyl residues are indicated by arrows.

(b) The dimeric structure is shown. This is very stable and probably represents the structure of Con A in the transglutaminase incubation mixtures.The residues belonging to the large β-structure at the back of the dimer are in bold type. These structures are taken from Reeke et al., (1975).

of six amino acid residues which are either surface located or fully exposed. By comparison with other substrate sites within globular proteins the size of this loop-like connecting section may be too small to allow reactivity. A similar sized loop is seen in ribonuclease A which also contains unreactive glutaminyl residues (see section 4.3.1.1.).

4.3.1.3. HORSE HEART CYTOCHROME C

A consideration of the X-ray structure of horse heart cytochrome c, (Bushnall et al., 1990) shows that Gln_8 is located within an α-helical region, that Gln_{12} is restrained by its interaction with the haem group, but that Gln_{41} is located in a surface loop and hence is the most exposed region of the protein (see Fig 4.4). This region has been shown to become highly flexible on acid denaturation even though secondary structure is retained in other parts of the molecule, (Jeng, 1990). The sequences surrounding the glutaminyl residues are shown in Fig 4.5.

```
          +   + +         +
1     K-G-K-K-I-F-V-Q-K-C-A-Q-C-H-T    15
              +               - +
5     I-F-V-Q-K-C-A-Q-C-H-T-V-E-K-G    19
            + +
36    L-F-G-R-K-T-G-Q-A-P-G-F-T-Y-T    50
```

Fig 4.5. THE SEQUENCES SURROUNDING GLUTAMINYL RESIDUES IN HORSE HEART CYTOCHROME C. [Positively charged residues are indicated with + symbols and negatively charged residues with a - symbol]

Although cytochrome c did not exhibit substrate properties for transglutaminase under the conditions used in Table 4.1, Butler and Landon, (1979) tested the substrate properties of ethanol denatured samples of cytochrome c and found that transglutaminase incorporated between 0.28 and 0.85 mol amine/mol protein. Limited studies on the modified proteins established that no radioactivity was associated with the haem-containing peptide, which contains glutaminyl residue Gln_{12}. Denatured protein was digested with trypsin and the peptides tested for substrate properties. In this case 0.91 mol amine/mol protein incorporation was recorded. Again no labelling of the haem peptide was observed. A consideration of the amino acid sequences surrounding the glutaminyl residues within cytochrome c indicates that although Gln_{43} is surrounded by the types of amino acid usually associated with the substrate sites of transglutaminase (eg Gly, Pro, and Thr residues) that Gln_{8}, and Gln_{12} are not. Hence Gln_{43} is likely to be the reactive site within cytochrome c in the denatured form and may account for the significant degree of modification by transglutaminase (ie. 0.2 mol amine/mol protein) observed in the native structure. This region is thought to become disordered under under mildly denaturing conditions (Jeng et al., 1990)) and this could account for the almost stoichiometric modification observed by Butler and Landon (1979) under such conditions.

The reason for the lack of reactivity of ethanol denatured cytochrome c in this report could be due to the fact that the ethanol was dialysed away proir to testing the substrate properties of the protein. Substantial refolding may have occurred during this time to give a fully or partially folded protein with similar substrate properties to the fully folded form.

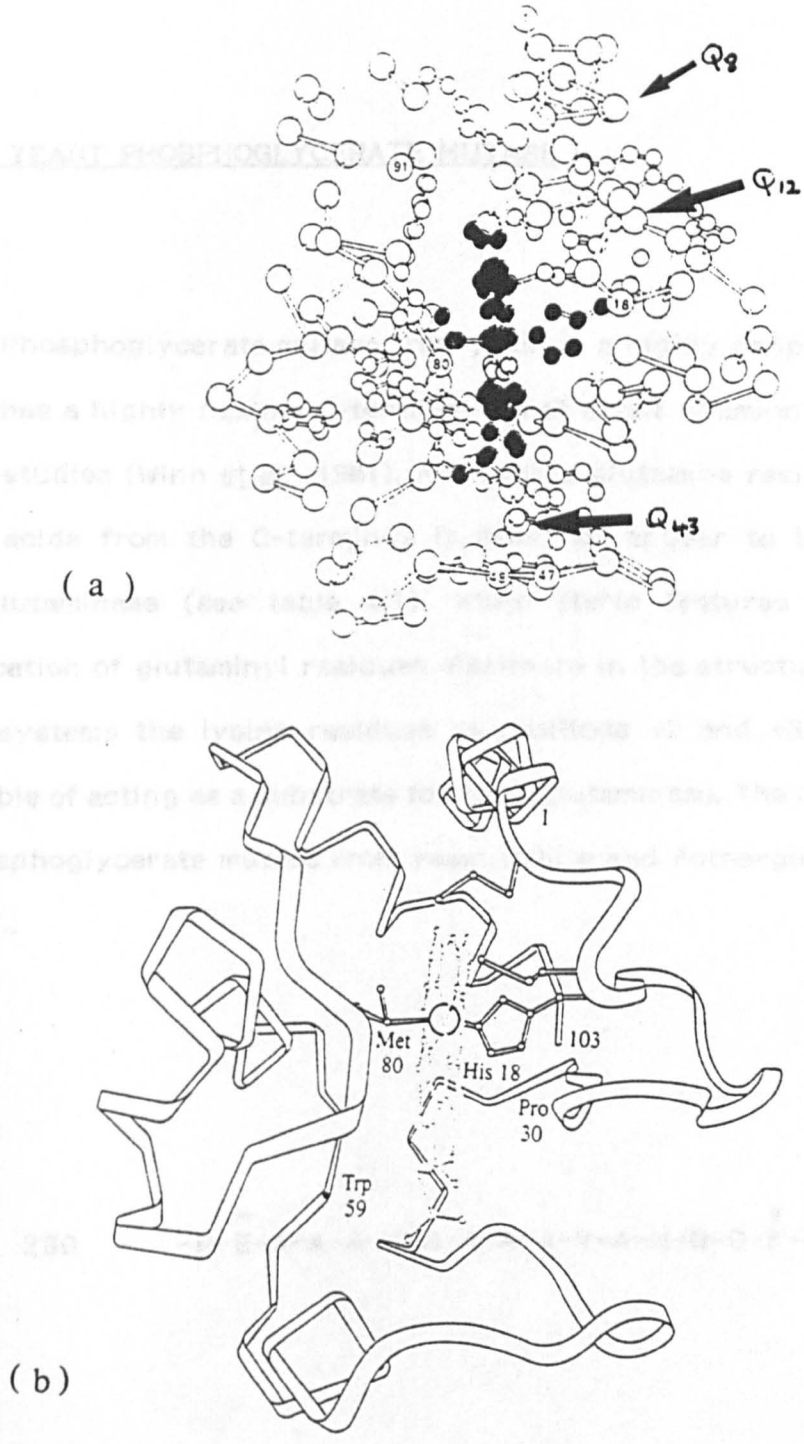


Fig 4.4. THE STRUCTURE OF HORSE HEART CYTOCHROME C

(a) Schematic drawing of the trace of the polypeptide in a subunit of cytochrome c from tuna. This protein is known to have a very similar tertiary folded structure to the protein purified from horse heart. The position of the glutaminyl residues within the structure are shown using arrows. The diagram is taken from "BIochemistry" (Rawn, 1989)

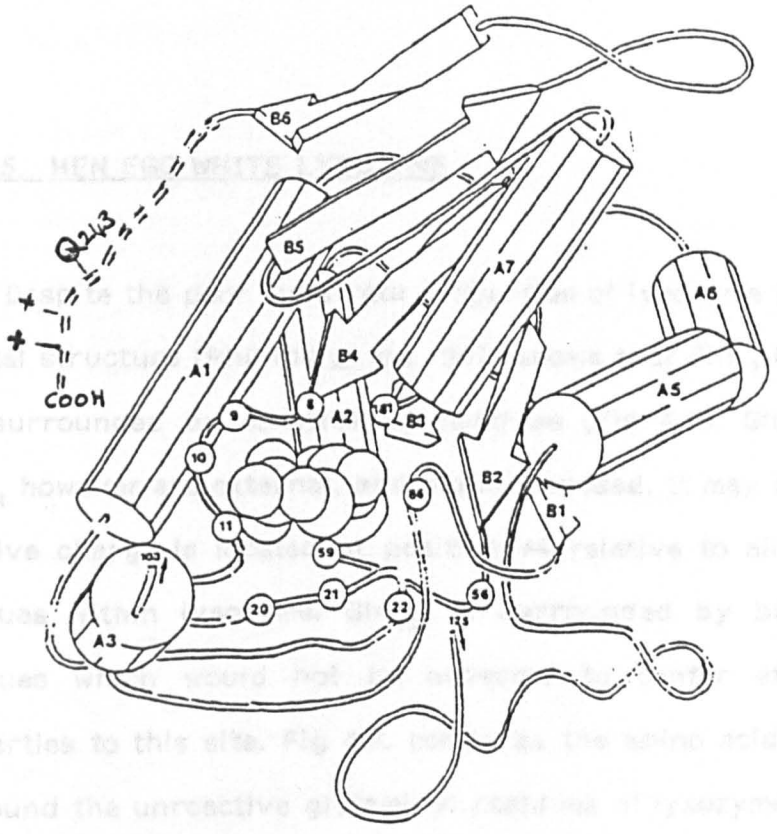
(b) The α-chain of cytochrome c. The diagram is taken from Dickerson, R.E. (1972). Some of the residues are numbered to guide the reader around the fold of the polypeptide chain. The protein sequence may be found in Appendix III.

4.3|4. YEAST PHOSPHOGLYCERATE MUTASE

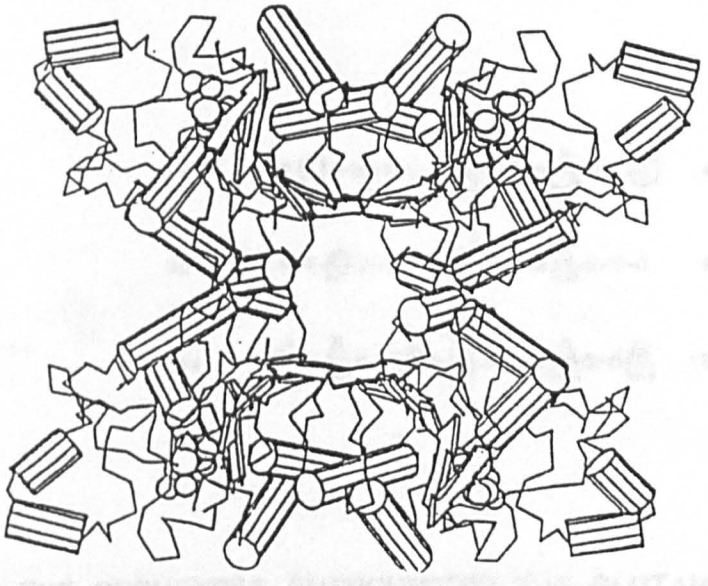
Phosphoglycerate mutase from yeast is a highly compact protein (Fig 4.6) which has a highly flexible C-terminal tail of about 10 amino acids, as judged by X-ray studies (Winn et al., 1981). Although a glutamine residue is located three amino acids from the C-terminus it does not appear to be reactive towards transglutaminase (see table 4.1). While steric features probably preclude modification of glutaminyl residues elsewhere in the structure, by analogy with other systems the lysine residues at positions +2 and +3 may render Gln_{243} incapable of acting as a substrate for transglutaminase. The C-terminal sequence of phosphoglycerate mutase from yeast (White and Fothergill, 1988) is shown in Fig 4.7..

230 -P-E-A-A-A-A-G-A-A-A-V-A-N-Q-G-K-K-COOH 246

Fig 4.7. THE SEQUENCE ASSOCIATED WITH THE C-TERMINAL "TAIL" OF YEAST PHOSPHOGLYCERATE MUTASE



(a)



(b)

Fig 4.6. THE STRUCTURE OF YEAST PHOSPHOGLYCERATE MUTASE

(a) Schematic drawing of the trace of the polypeptide in a subunit of phosphoglycerate mutase from yeast. A portion of the polypeptide chain has been removed to reveal the α-carbon positions occupied by bound ligands. Regions of ß-sheet are arrow-like bars and α-helix are tubes. The last fourteen residues at the carboxyl terminus were not observed in the crystal structure and are thought to form a flexible tail (see text).

(b) A representation of the tetrameric structure of PGM.

4.3.1.5. HEN EGG WHITE LYSOZYME

Despite the poor substrate properties of lysosyme (see Table 4.1) the crystal structure (Phillips et al., 1967) shows that Gln_{57} is surface exposed but surrounded by constrained residues (Fig 4.8). Glutamine Gln_{41} and Gln_{121} however are external, and highly exposed. It may be significant that positive charge is located at position +4 relative to all three glutaminyl residues within lysozyme. Gln_{121} is surrounded by bulky hydrophobic residues which would not be expected to confer efficient substrate properties to this site. Fig 4.9. compares the amino acid sequences which surround the unreactive glutaminyl residues of lysozyme.

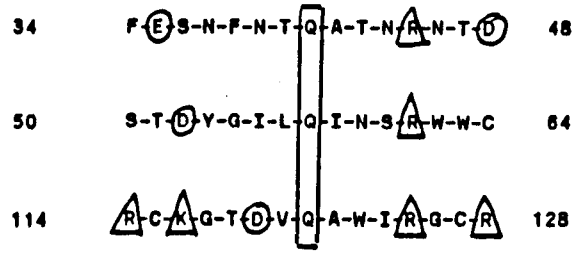
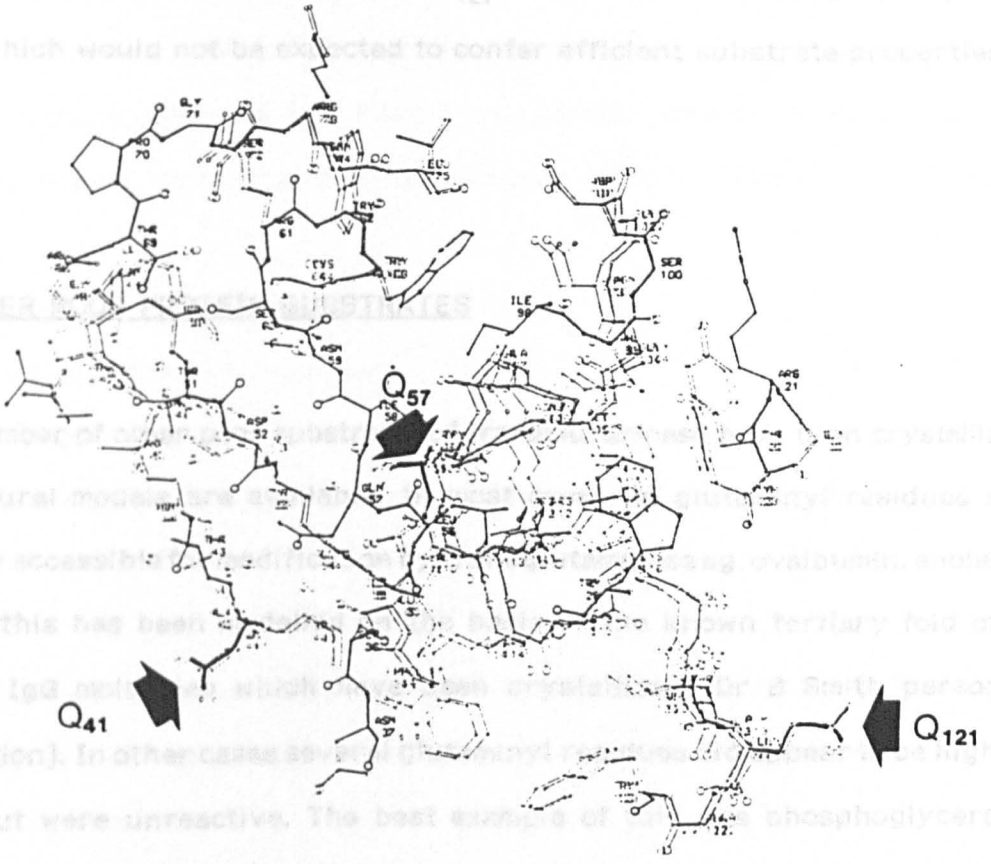


Fig 4.9. THE SEQUENCES SURROUNDING THE GLUTAMINYL RESIDUES OF LYSOSYME positively charged residues are in triangles, negatively charged residues are in circles

	1				5				10				15	
1	LYS	VAL	PHE	GLY	ARG	CYS	GLU	LEU	ALA	ALA	ALA	MET	LYS	ARG	HIS	15
16	GLY	LEU	ASP	ASN	TYR	ARG	GLY	TYR	SER	LEU	GLY	ASN	TRP	VAL	CYS	30
31	ALA	ALA	LYS	PHE	GLU	SER	ASN	PHE	ASN	THR	GLN	ALA	THR	ASN	ARG	45
46	ASN	THR	ASP	GLY	SER	THR	ASP	TYR	GLY	ILE	LEU	GLN	ILE	ASN	SER	60
61	ARG	TRP	TRP	CYS	ASP	ASN	GLY	ARG	THR	PRO	GLY	SER	ARG	ASN	LEU	75
76	CYS	ASN	ILE	PRO	CYS	SER	ALA	LEU	LEU	SER	SER	ASP	ILE	THR	ALA	90
91	SER	VAL	ASN	CYS	ALA	LYS	LYS	ILE	VAL	SER	ASP	GLY	ASP	GLY	MET	105
106	ASN	ALA	TRP	VAL	ALA	TRP	ARG	ASN	ARG	CYS	LYS	GLY	THR	ASP	VAL	120
121	GLN	ALA	TRP	ILE	ARG	GLY	CYS	ARG	LEU							135

(a)



(b)

Fig 4.8 THE STRUCTURE OF HEN EGG WHITE LYSOZYME

(a) Amino acid sequence of hen egg white lysosome (Canfield, 1963).

(b) A stereoscopic high resolution model of lysosome based on the crystal structure determined by Phillips (1967). The highly exposed glutaminyl residues are indicated by arrows. The large hydrophobic Trp_{123} does not appear to block Gln_{121} and appears to be on a particually exposed C-terminal arm of the protein. Its lack of reactivity may be due to chemical features within the linear sequence surrounding this residue. Glutaminyl residue Gln_{57} is probably more constrained in the regions either side of this residue due to the ß-sheet interactions within the main chain at this point.

Although the native protein has poor substrate properties, Yan and Wold, (1984) have reported 1.0 mol amine/mol protein incorporation of dansylcadaverine into the succinylated and reduced (ie partially denatured) protein. Inspection of the sequences which surround the glutaminyl side-chains of lysozyme, only Gln_{41} contains sequence motifs seen in other transglutaminase substrates ie. it is surrounded by threonine side-chains. Gln_{121} is surrounded by bulky hydrophobic residues which would not be expected to confer efficient substrate properties to this site.

4.3|6. OTHER POOR PROTEIN SUBSTRATES

A number of other poor substrates of transglutaminase have been crystallized and structural models are available. In most cases no glutaminyl residues are sufficiently accessible for modification by transglutaminase eg. ovalbumin, enolase, IgG B72.3 [this has been modelled on the basis of the known tertiary fold of a number of IgG molecules which have been crystallized (Dr B Smith personal communication]. In other cases several glutaminyl residues did appear to be highly exposed, but were unreactive. The best example of this was phosphoglycerate kinase from yeast. The lack of reactivity was presumed to be due to the presence of neighbouring positively charged residues and this was investigated in greater depth by testing the substrate properties of a number of mutant forms of the enzyme (see chapter 6).

4.4. GLOBULAR PROTEIN SUBSTRATES WITH SUBSTRATE PROPERTIES FOR TRANSGLUTAMINASE

4.4.1. INTRODUCTION

Globular proteins which have both substrate properties for transglutaminase, and have X-ray crystal structures available tend to have highly exposed glutaminyl residues either within:

a) extended N- or C- termini,

b) large (probably) flexible surface loops,

c) have sequences which appear to favour β-turn structure.

These features are associated with a number of the protein substrates listed in Table 4.2. The structures of two such proteins are discussed in sections 4.4.2. and 4.4.3.

4.4.2. RABBIT MUSCLE ALDOLASE

Aldolase (rabbit muscle) was modified up to a stoichiometry of between 1.0 and 2.0 mol/mol. This is a comparable figure to that obtained by Brenner and Wold (1978). Inspection of the X-ray structure determined to 2.7 Å (Sygusch et al.,1987) shows that the enzyme is folded into a single highly compact domain structure which is shown in Fig 4.10. The protein appears

to favour tetrameric quaternary structure and this is stabilized by interactions between the flanking α-helices of the ß-barrel. Hence glutaminyl side-chains at positions Gln_{200}, Gln_{204}, Gln_{244}, Gln_{245}, Gln_{239}, Gln_{222}, are unlikely to be sufficiently exposed to be modified. Other glutamines which appear to be located within highly folded regions include: Gln_{10}, Gln_{34}, Gln_{44}, Gln_{60}, Gln_{85}, Gln_{95}, Gln_{136}, Gln_{178}, Gln_{179}, Gln_{304}, Gln_{322}, and Gln_{337}. The remaining glutaminyl side chains appear to be highly exposed.

Glutaminyl residues Gln_{125} is located in a flexible twenty two residue loop, Gln_{155} is located on a smaller eight residue surface loop and Gln_{345} is located in the electron density deficient C-terminal region of aldolase. The sequences (Tolan et al., 1984) associated with the exposed glutaminyl residues within rabbit muscle aldolase are shown in Fig 4.11.

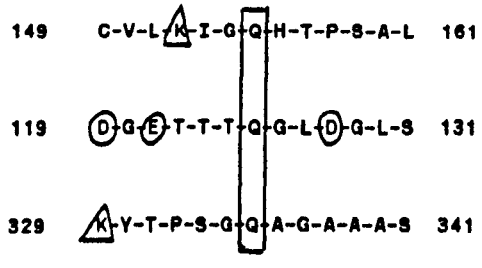


Fig 4.11. THE SEQUENCES WHICH SURROUND EXPOSED GLUTAMINYL RESIDUES IN ALDOLASE positively charged residues are in triangles and negatively charged residues are in circles

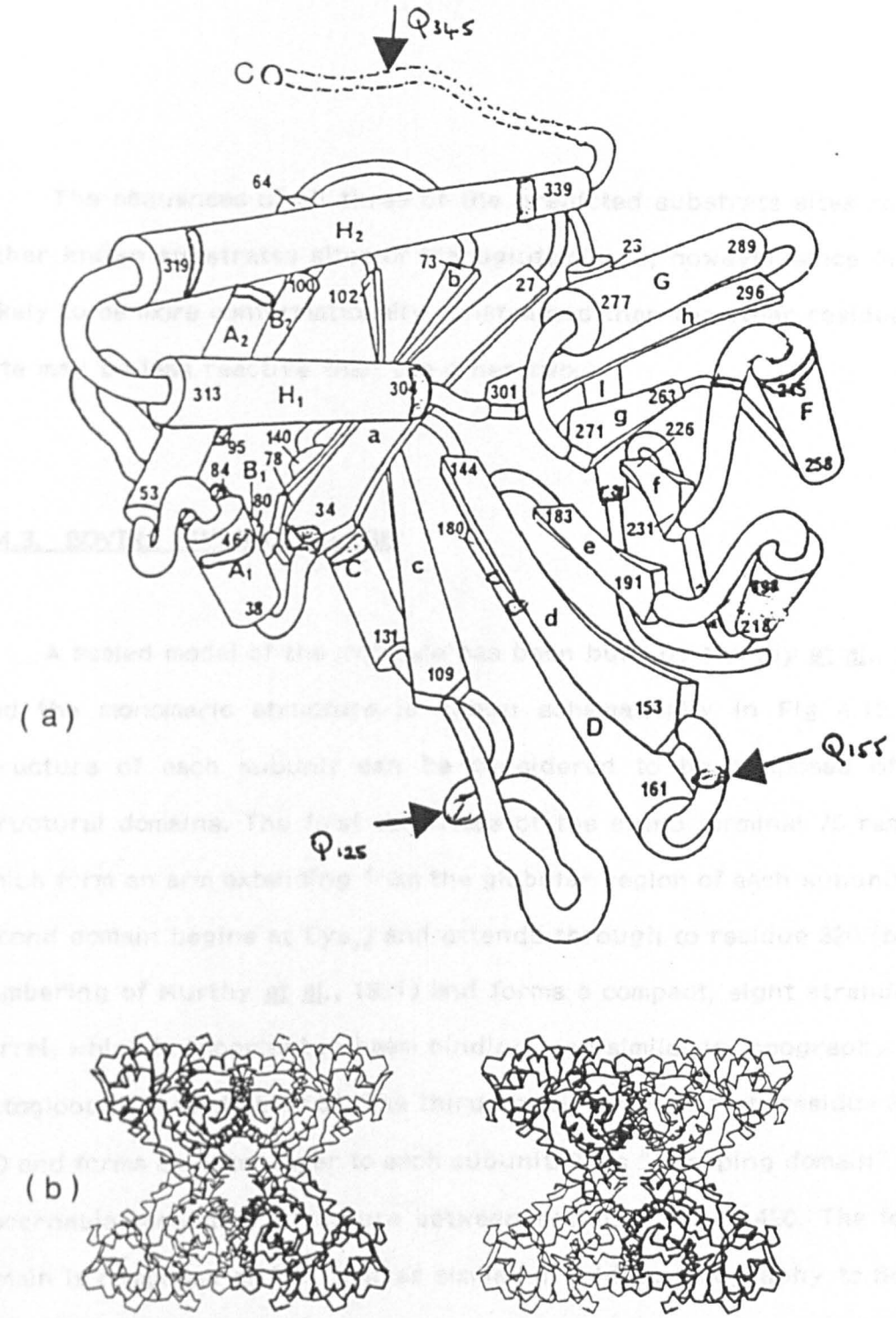


Fig 4.10. THE STRUCTURE OF RABBIT MUSCLE ALDOLASE

(a) Schematic drawing of the trace of the polypeptide in a subunit of rabbit skeletal muscle aldolase. Regions of ß-sheet are arrow-like bars and ∝-helix are tubes. The numbering represents the first and last residues of the amino acid sequence corresponding to each region. The approximate location of potentially reactive glutaminyl residues are indicated by arrows. The carboxyl terminus of the protein has not been identified unambiguously (Sygusch et al., 1987).

(b) The ∝-carbon backbone representation of the tetramer.

The sequences of all three of the predicted substrate sites resemble other known substrates sites of transglutaminase, however since Gln_{155} is likely to be more conformationally constrained than the other residues this site may be less reactive than the other two.

4.4.3. BOVINE LIVER CATALASE

A scaled model of the molecule has been built by Murthy et al., (1981) and the monomeric structure is shown schematically in Fig 4.12.. The structure of each subunit can be considered to be composed of four structural domains. The first comprises of the amino terminal 75 residues which form an arm extending from the globular region of each subunit. The second domain begins at Lys_{76} and extends through to residue 320 (by the numbering of Murthy et al., 1981) and forms a compact, eight stranded β-barrel, which is important in haem binding, and similar in topography to β-lactoglobulin (see chapter 5). The third domain extends from residue 321 to 420 and forms an outer layer to each subunit. This "wrapping domain" lacks discernable secondary structure between residues 366 to 420. The fourth domain is composed of four helices similar in folding topography to helices E,F,G,H, in haemoglobin (see Appendix I vii).

By comparing the amino acid sequence of catalase (Schroeder et al., 1982) with the crystal structure, and using the criteria of Chothia (1975), residues Gln_{17}, Gln_{52}, Gln_{386}, Gln_{414}, Gln_{493} appear to be the most highly exposed. Glutaminyl residues $Gln_{,397}$, and Gln_{429} may also be exposed. A consideration of the amino acid residues surrounding these residues indicates that residues Gln_{17}, Gln_{386}, Gln_{429}, and Gln_{493} have positively charged residues at positions within five residues to the C- terminal side

of them within the protein sequence (Schroeder et al., 1982). The sequences surrounding the remaining glutaminyl residues are shown in Fig 4.13.

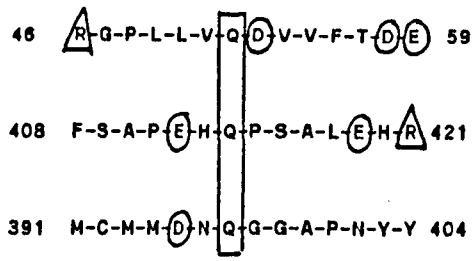


Fig 4.13. THE SEQUENCES SURROUNDING SOME HIGHLY EXPOSED GLUTAMINYL RESIDUES IN BOVINE LIVER CATALASE positively charged residues are in triangles and negatively charged residues are in circles

The sequences which surround glutaminyl residues Gln_{52} and Gln_{397} contain features which resemble those in other transglutaminase substrates. Glutaminyl residue Gln_{413} has a neighbouring C-terminal proline residue and this is not normally associated with favourable substrate properties. Glutaminyl residue Gln_{413} is therefore predicted to be a poorer substrate relative to the other two. On the basis of further indirect evidence the substrate sites within bovine liver catalase were further characterized by peptide mapping, see section 8.2..

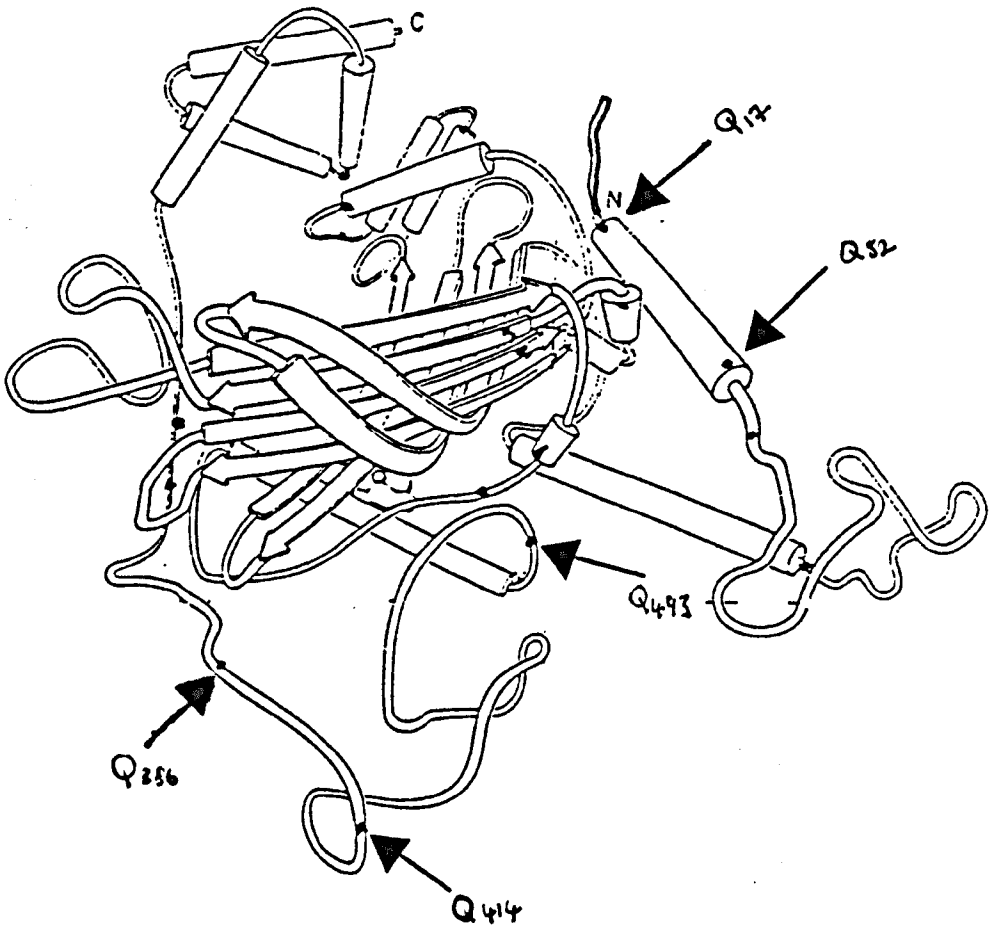


Fig 4.12. THE STRUCTURE OF BOVINE LIVER CATALASE

Schematic drawing of the trace of the polypeptide in a subunit of bovine liver catalase. Regions of β-sheet are arrow-like bars and α-helix are tubes. The location of exposed. The approximate location of surface exposed (and therefore potentially reactive) glutaminyl residues are indicated with arrows. The protein has been crystallized as tetramer (Eventoff et al., 1976) but under the conditions of labelling is likely to resemble the monomer (see text). The diagram of the monomer is from Murthy et al., 1981)

4.4. THE DIRECT DETERMINATION OF REACTIVE GLUTAMINYL RESIDUES WITHIN GLOBULAR PROTEINS

The poor substrate properties of the globular substrates tested could generally be explained on the basis of the apparent determinants of specificity of transglutaminase discussed in section 1.7., particularly that associated with C-terminal positive charge. In order to further test this hypothesis some of the protein substrates listed in Table 4.2 were further characterized by the determination of the location of reactive and non-reactive glutaminyl residues within their structures. This was done in the case of: bovine β-lactoglobulin (see chapter 5), the yeast phosphoglycerate kinase (PGK) $His_{388} \rightarrow Gln_{388}$ mutant protein (see chapter 6), non-succinylated (native) bovine β-casein (see chapter 7), and porcine pepsin (see chapter 8).

The position of the reactive glutaminyl residues within β-lactoglobulin (see section 5.3.) were correctly predicted on the basis of comparisons with other known substrate sites (see section 1.7.) in a preliminary communication (Coussons et al., 1990), prior to their direct determination, [Coussons et al., (1991); submitted for publication].

CHAPTER FIVE: AN INVESTIGATION OF THE SUBSTRATE PROPERTIES OF β-LACTOGLOBULIN (BOVINE) FOR GUINEA PIG LIVER TRANSGLUTAMINASE

5.1. INTRODUCTION

The substrate properties of β-lactoglobulin for transglutaminase were investigated and the specificity of transglutaminase was assessed on the basis of its selectivity towards glutaminyl residues within this protein. Bovine β-lactoglobulin exists as two major isoforms. These A and B variants have identical amino acid sequences except for a $Val_{68} \rightarrow Ala_{68}$, and an $Asp_{122} \rightarrow Gly_{122}$ mutation in isoform B (reviewed by Lyster, 1972). In the experiments described, no detectable differences were observed between the behaviour of the A and B forms or the mixture, the data presented refers to the mixture.

METHODS

The buffers used for the modification of β-lactoglobulin by transglutaminase and that used for recording the c.d. spectrum of the protein were slightly different in composition. The modification buffer contained a more concentrated (0.1 M) triethanolamine buffer (together with standard reaction mixture components, see section 2.2.3.2.), while spectra were recorded in either 20 mM triethanolamine or 20 mM Tris-MES buffers. This was necessary due to the high absorbance of the buffer used for modification below 210 nm and the requirement for buffering capacity over the range from pH 5.5 to pH 9.0. The concentration of β-lactoglobulin was determined spectrophotometrically at 278 nm using an absorption coefficient of 0.96 ml. $mg.^{-1}.\ cm^{-1}$ (Kella and Kinsella, 1988).

C.D. measurements were performed twice and traces are the average of two scans. Labelling experiments were performed at least twice in duplicate.

5.2. RESULTS AND DISCUSSION

5.2.1. SOLUTION STRUCTURE OF BOVINE β-LACTOGLOBULIN

β-lactoglobulin appeared to retain a considerable amount of structure at pH values up to 8.0 even in the presence of DTT. Analysis of the far u.v. spectrum obtained at pH 7.6 (Fig 5.1a) from 240 nm to 190 nm, according to the CONTIN procedure (Provencher and Glöckner, 1981) [see section 2.2.8.3.] gave the following estimates of secondary structure; 17 % (+/- 1.1 %)α-helix , 45 % (+/- 1.5 %) β-sheet, 38 % (+/- 1.8 %) remainder. These values are similar to those of an earlier c.d. study [17 %α-helix and 41 % β-sheet (Takeda and Moriyama, 1989)]. The results of X-ray crystallography (Papiz et al., 1986) indicates that there are 11 and 83 out of the total of 162 amino acids in α-helix and β-sheet respectively (ie. 7 % and 51 % respectively). It is of interest that using the method of Siegel et al., (1980) which analyses the c.d. spectrum over a smaller range of wavelengths (240 to 210 nm), the α-helix content is estimated to be 11 (+/- 2 %), which is closer to the value obtained by X-ray crystallography.

The far u.v. c.d. spectra of bovine β-lactoglobulin at pH values of 5.5, 7.6, and 9.0 in 20 mM MES plus 20 mM Tris buffer are shown in Fig 5.1a. There are no significant differences between the spectra obtained in the two different buffers at pH 7.6. There are small differences in the spectra obtained at pH 5.5 and pH 7.6, which reflect a small decrease in α-helix content (Siegel et al., 1980) from 13 % at pH 5.5 to 11 % at pH 7.6. The α-helix content at pH 9.0 is 10 %. The structural changes occurring between

pH 5.5 and 7.6 correspond to the well documented Tanford transition (Lyster, 1972).

In the near u.v. (Fig 5.1c) the c.d. spectra show signals in the 280 to 300 nm region characteristic of Trp and Tyr side chains. The diminution in signal between pH 5.5 and 7.6 reflect changes in the secondary structure which affect the environment of aromatic side chains. The Tanford transition has been previously proposed to involve changes in the environment of one Tyr side chain (Lyster, 1972). When bovine ß-lactoglobulin was incubated at pH 5.5 and pH 7.6 for 24 hr at 25°C there were no significant changes in either the near or far u.v. spectrum. At pH 9.0 the amplitude of the near u.v. spectrum decreased approximately by 20 % over this time, suggesting some loss of native tertiary structure. Instability of the protein above pH 8.0 has been previously commented on (Lyster, 1972).

The spectra of ß-lactoglobulin in the presence of 1.0 mM DTT were obtained over the same pH range. The far u.v. spectra indicated that immediately after addition of reducing agent there are only small changes in the secondary structure of the protein (data not shown). In the near u.v., the amplitude of the spectrum at pH 5.5 is about 20 % less than in the absence of DTT. This effect did not occur at pH 7.6 or pH 9.0. However at pH 9.0 there are considerable time dependent structural changes observed in both the far and near u.v. spectra (Figs 5.1d and 5.1e). Such effects were not observed at the lower pH values tested.

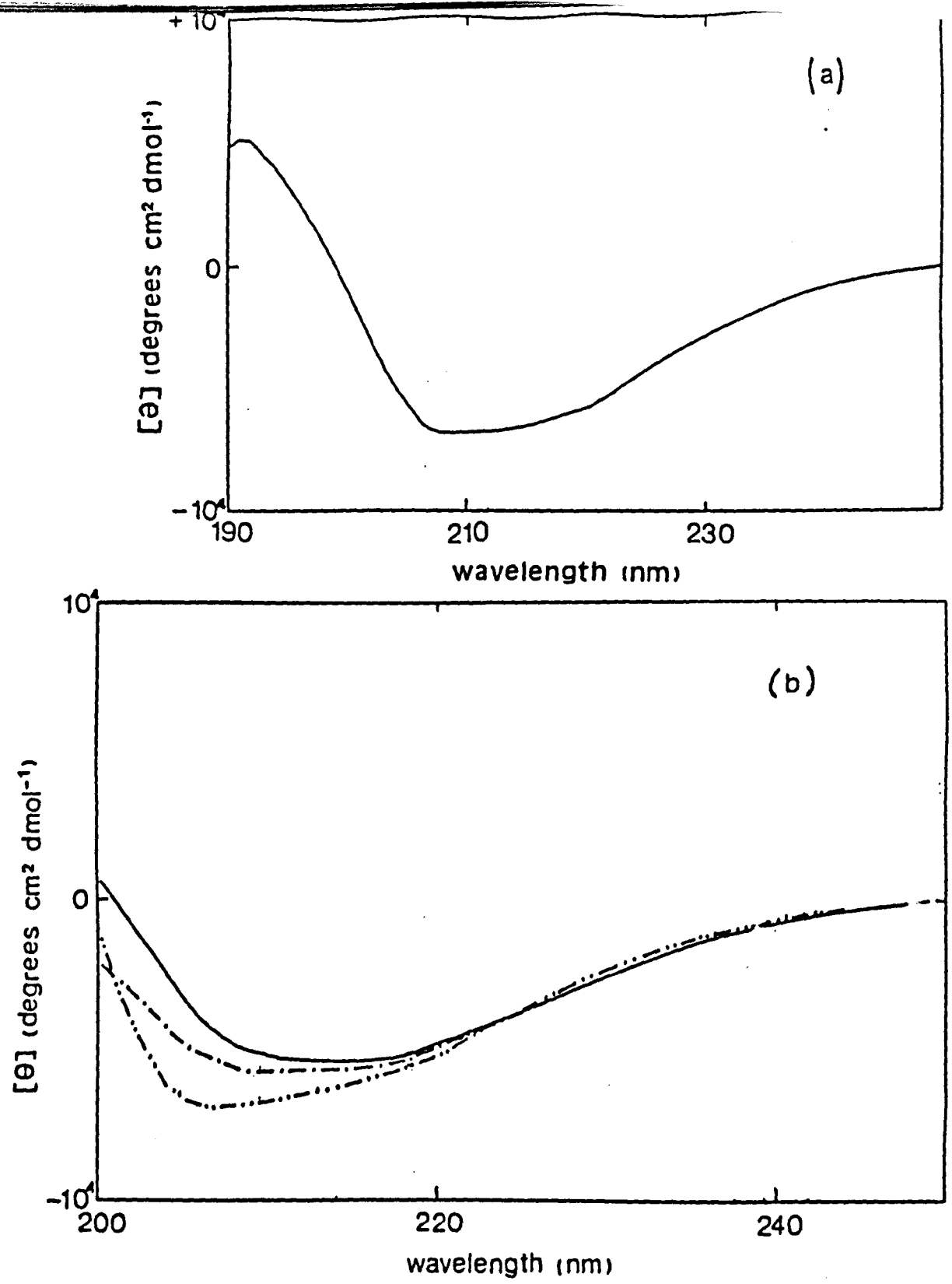


Fig 5.1. C.D SPECTRA OF β-LACTOGLOBULIN

Spectra were recorded at 25 °C at a protein concentration of 0.9 mg/ml in cells of pathlength of 0.02 cm (far u.v.) or 0.5 cm (near u.v.).

(a) Far u.v. spectrum in 20 mM triethanolamine-HCl buffer, pH 7.6.
(b) Far u.v. spectra in 50 mM tris buffers. (-.-) . pH 5.5; (—) , pH 7.5 ; (- . . -), pH 9.0. These traces were not affected by the presence of 50 mM MES in the incubation mixture.

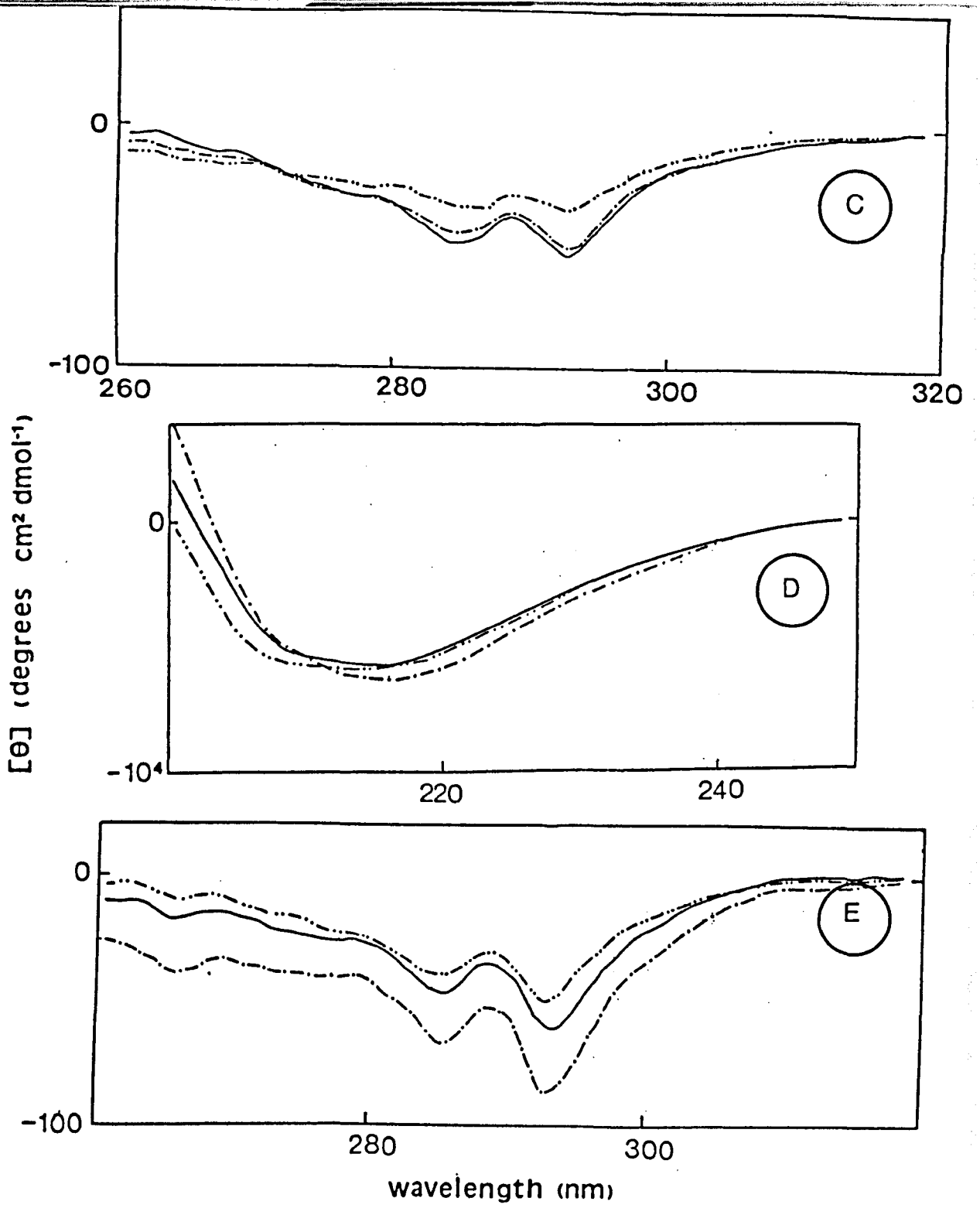


Fig 5.1. (cont.) C.D. SPECTRA OF β-LACTOGLOBULIN

(c) Near u.v. spectra in the same buffers as (b). The spectrum at pH 7.6 is identical to that recorded in 20 mM Triethanolamine-HCl buffer at this pH.

(d) Far u.v. spectra in 50 mM MES plus 50 mM Tris buffer , pH 9.0, in the presence of 1 mM dithiothreitol. (——).

(- . -) and (- . . -) refer to spectra recorded after incubation for 1 min , 4 hours and 24 hours respectively.

(e) Near u.v. spectra recorded under the same conditions as in (d).

5.2.2. THE TRANSGLUTAMINASE-CATALYSED MODIFICATION OF β-LACTOGLOBULIN

The time course of incorporation of putrescine into β-lactoglobulin over a range of pH is shown in Fig 5.2. The maximum incorporation observed was 3.0 mol putrescine/mol β-lactoglobulin monomer at pH 9.0. Approximately 2.0 mol putrescine/mol protein monomer was incorporated at pH 7.6 ie. the pH at which the crystal structure (Papiz et al., 1986) was determined. When dansylcadaverine was substituted for putrescine in incubation mixtures only 1.2 mol dansylcadaverine/mol protein monomer was incorporated after 20 hr. This probably reflects steric blocking of a second reactive site by the bulky dansyl group of dansylcadaverine, following covalent attachment of the first molecule to a more reactive site. Labelling was only observed in the presence of transglutaminase and Ca^{2+}. The presence of DTT was required for protein modification, since in its absence the stoichiometry of labelling was reduced to < 0.2 mol amine/mol (see Fig 5.3).

Bovine β-lactoglobulin has been shown to exist in solution as an equilibrium mixture of different conformations and quaternary structures. The effect of pH on bovine β-lactoglobulin structure has been studied (McKenzie and Sawyer, 1967). Above pH 8.0 the predominant form is thought to be the monomer. This form may be unstable at high pH and prone to denaturation. This process may expose further site(s) for transglutaminase catalysed modification, which is/are not normally accessible at pH 7.6. This may explain the increase in maximum observed labelling stoichiometry to 3.0 mol putrescine/mol protein monomer at pH 9.0. From SDS-PAGE analysis (Laemmli, 1970) of the modified protein on 10% acrylamide gels, the was no evidence of transglutaminase catalysed cross-linking of β-lactoglobulin.

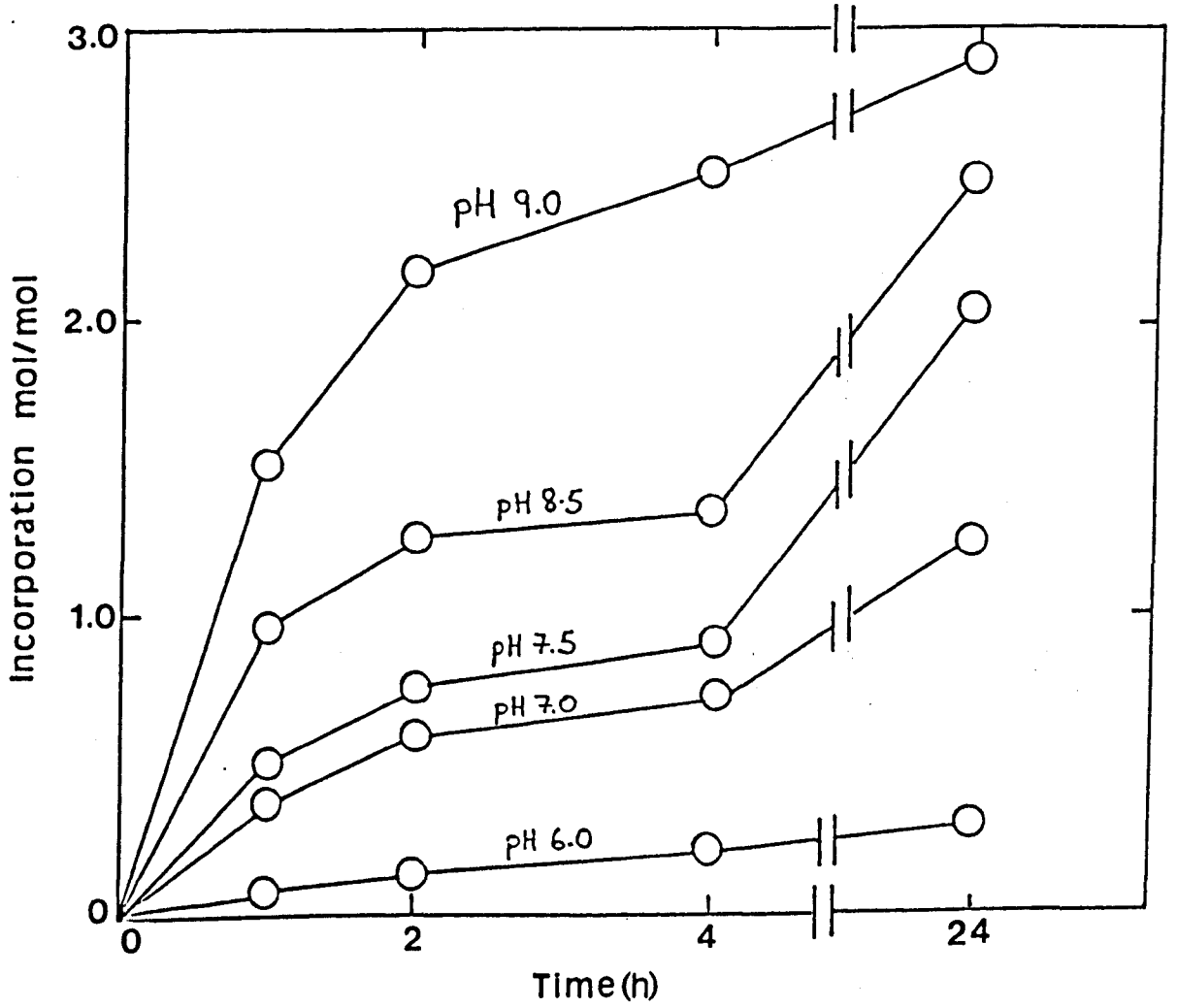


Fig 5.2. THE EFFECT OF pH ON THE INCORPORATION OF PUTRESCINE INTO BOVINE β-LACTOGLOBULIN

Protein concentration was 2.0 mg/ml and reaction was carried out at 20°C. DTT (1.0 mM) was included in the reaction mixture. No labelling of β-lactoglobulin was observed in the absence of transglutaminase. The buffer system is described in the text.

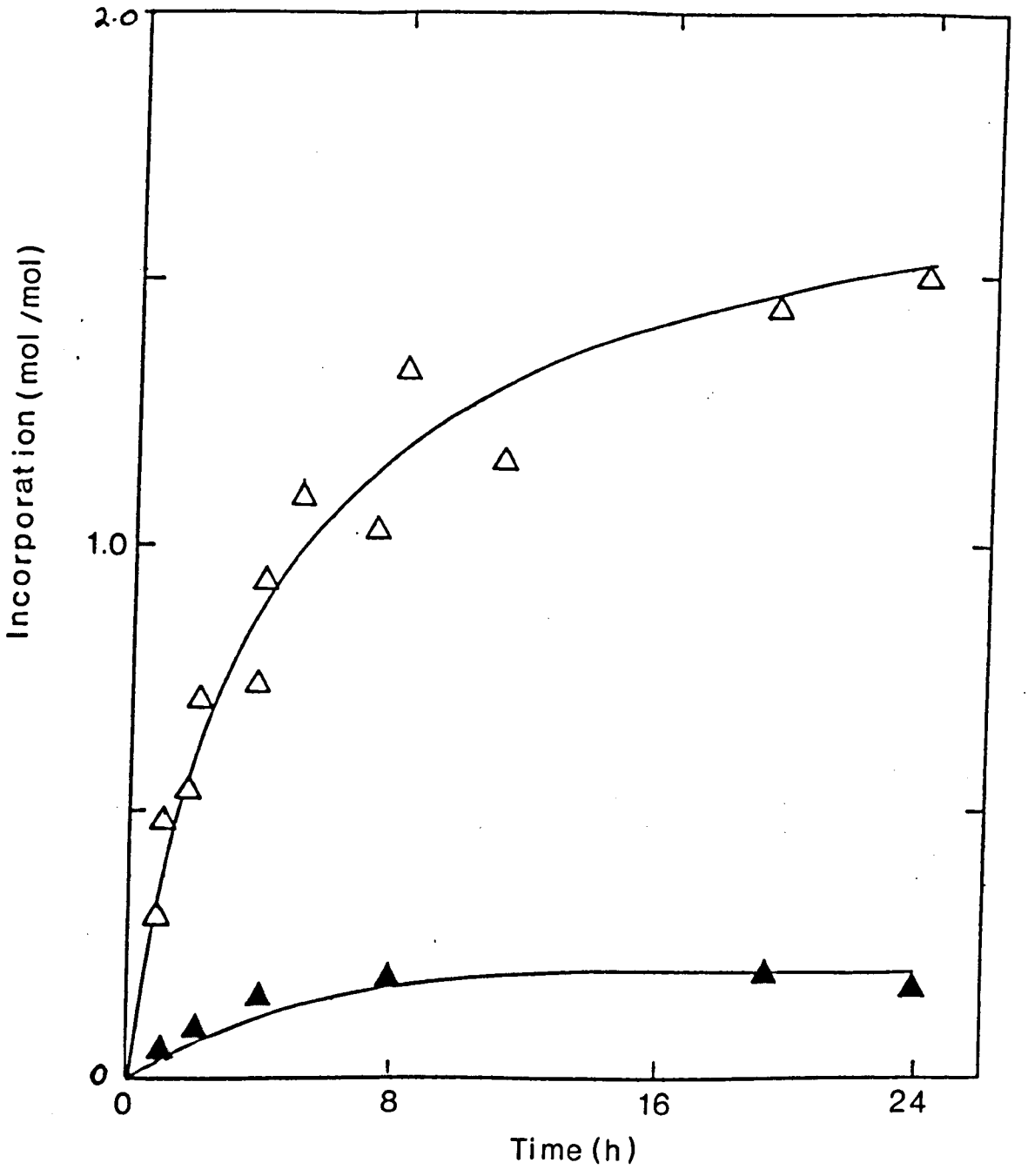


Fig 5.3 THE MODIFICATION OF β-LACTOGLOBULIN BY TRANSGLUTAMINASE

The time course is shown of the transglutaminase-catalysed incorporation of [^{14}C] putrescine into a mixture of the A and B forms of bovine β-lactoglobulin. The reaction was carried out at 25°C in 0.1 M Triethanolamine-HCl buffer at pH 7.6 as described in the methods section. Open and closed symbols refer to reactions carried out in the absence and presence of 1 mM DTT respectively.

5.2.3. pH DEPENDENT STRUCTURAL TRANSITIONS IN β-LACTOGLOBULIN DO NOT AFFECT ITS SUBSTRATE PROPERTIES

The labelling sites within bovine β-lactoglobulin were determined at pH 7.6 ie. that at which the crystal structure was determined (Papiz et al., 1986). The pH range covered included those around which bovine β-lactoglobulin undergoes the Tanford transition (Lyster 1972). The conformational change associated with this transition did not appear to directly affect the substrate properties of β-lactoglobulin, since when the initial rate of amine incorporation into the protein at each pH was divided by the initial rate of labelling of the synthetic peptide CBZ-Gln-Gly, the increase in the initial rate of labelling due to pH dependent conformational change in bovine β-lactoglobulin was seen to increase continuously over the pH range examined, (see Fig 5.4). If the conformational change associated with the Tanford transition had directly affected the reactivity of glutaminyl residues within the protein then the ratio of initial rate of protein modification/initial rate of CBZ-Gln-Gly modification would be expected to change in a non-linear way as the pH approached 7.6.

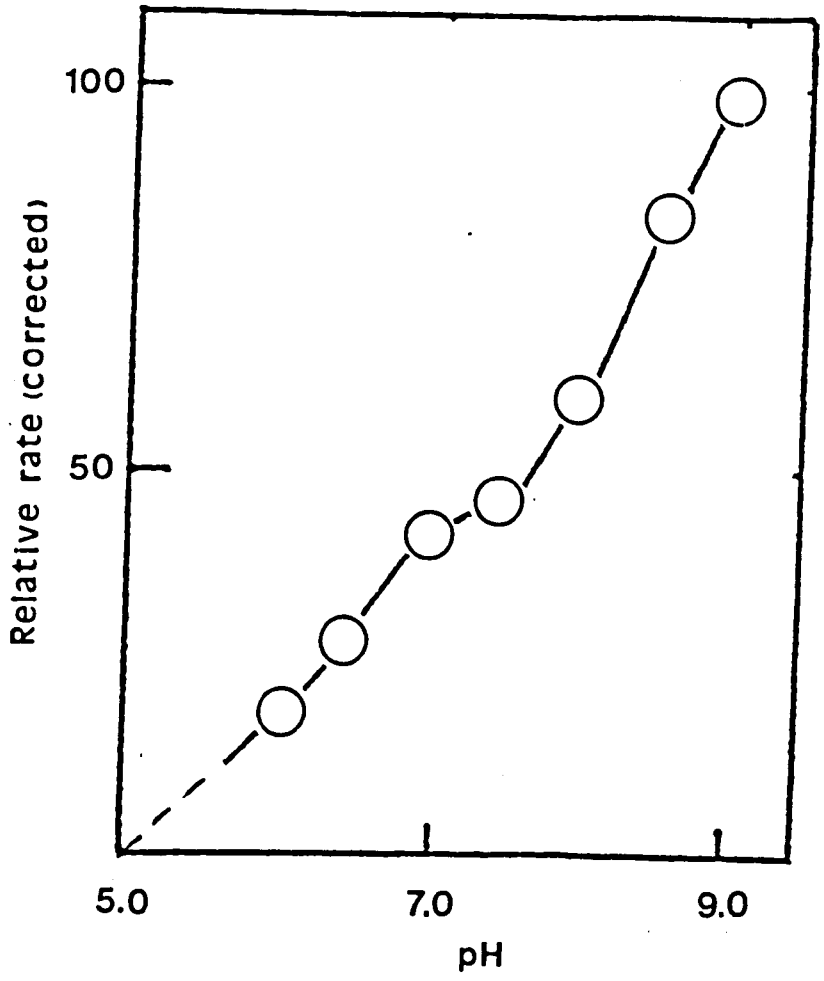


Fig 5.4. pH DEPENDENT STRUCTURAL CHANGE IN BOVINE β-LACTOGLOBULIN

The initial rates of putrescine incorporation into β-lactoglobulin were calculated from incorporation curves such as that shown in Fig 5.5 and these values were divided by the rate of hydroxylamine incorporation into CBZ-Gln-Gly using the method described in section 2.2.3.1. over the same time period. Since the pH is not expected to affect the conformation of CBZ-Gln-Gly or the charged state of the major charged amino acid side chains then any change in the ratio of the rate of reaction of transglutaminase towards these substrates is taken to be due to conformational change within β-lactoglobulin. Reactions were carried out with a lactoglobulin concentration of 2.0 mg/ml at 25°C over a pH range from pH 5.5 to pH 9.0.

5.3. IDENTIFICATION OF REACTIVE GLUTAMINYL RESIDUES FOR TRANSGLUTAMINASE WITHIN BOVINE β-LACTOGLOBULIN

5.3.1. LOCATION OF MODIFIED SITES WITHIN THE PRIMARY STRUCTURE

β-lactoglobulin which had been reacted for 20 hr in the presence of 1 mM DTT at pH 7.6 with either with dansylcadaverine, or putrescine was recovered by gel filtration and subjected to reduction, carboxymethylation and tryptic digestion (see section 2.2.5.). Modified peptides were purified by HPLC and detected by their fluorescence or radioactivity (see section 2.2.3.). The HPLC profiles of the products of digestion of both unmodified and dansylated β-lactoglobulin are shown in Figs 5.5 a) and b). The modified peptides which corresponded to the two fluorescent peaks at 42 % and 48 % solvent B (isopropanol) were collected and rechromatographed as described elsewhere (2.2.5.5.). The sequences of the modified peptides are shown in Fig 5.6. Portions of the PTH derivatives obtained at each cycle were analyzed for radioactivity or fluorescence (Coussons et. al., 1991) in order to confirm that any unidentified derivative corresponded to a site of incorporation of putrescine or dansylcadaverine.

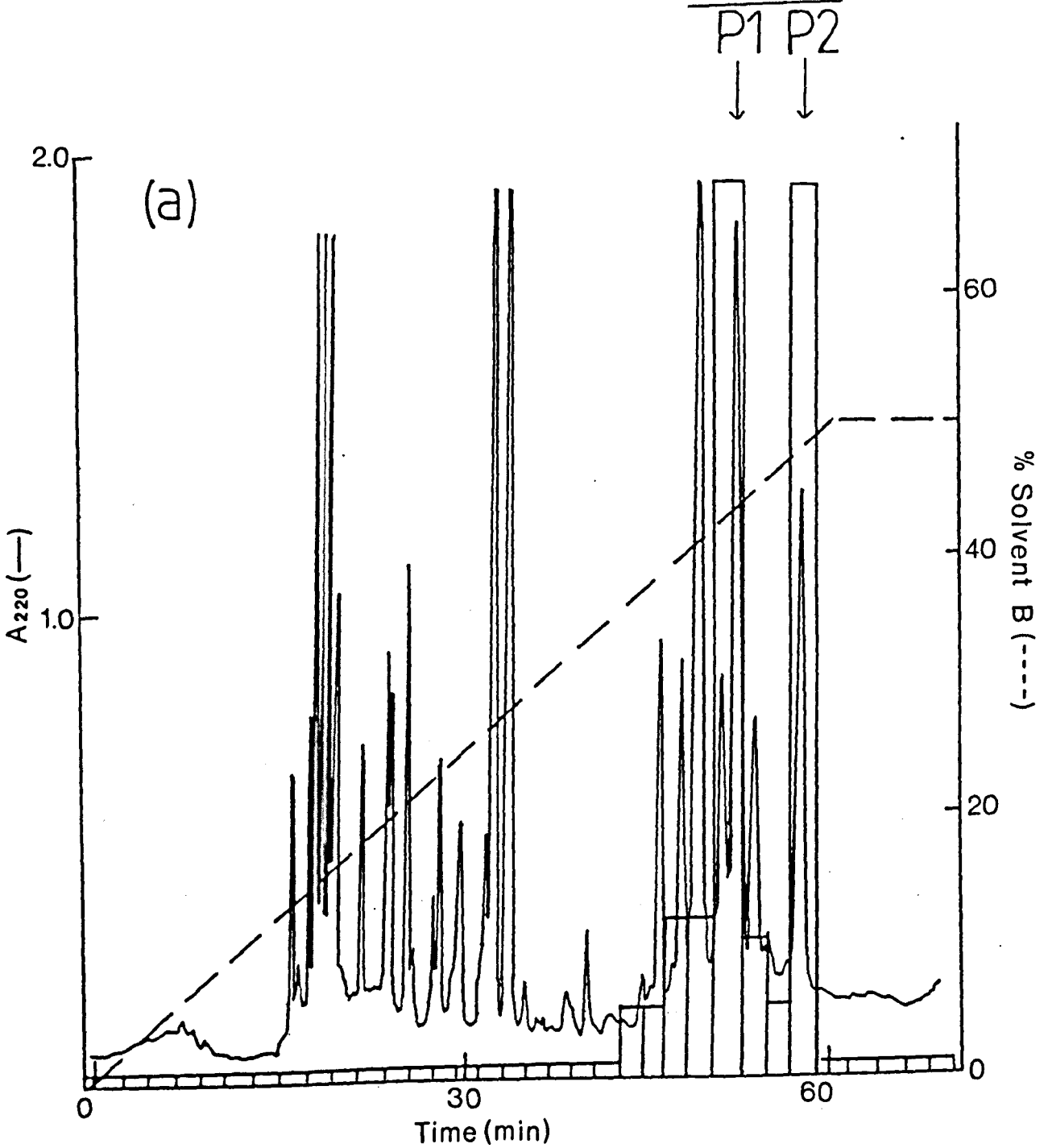


Fig 5.5.(a) PURIFICATION BY H.P.L.C. OF PEPTIDES PRODUCED BY TRYPTIC DIGESTION OF β-LACTOGLOBULIN

The modification of β-lactoglobulin was performed as described in the experimental section. Both traces (a) and (b) refer to separations of peptides on a Vydac C-18 column for dansylcadaverine modified and unmodified protein respectively. In each case approximately 0.2 mg of digested protein was applied. Solvent A consisted of aqueous 0.4% triethylamine, adjusted to pH 2.5 with orthophosphoric acid. Solvent B was solvent A containing 60 % (v/v) propan-2-ol. In (a) the rectangles show the relative fluorescence of the fractions (excitation and emission wavelengths 320 and 520 nm respectively). The relative amounts of P1 and P2, as judged by measurements of A_{220} were 1.4 : 1.0. Samples were further purified as described in section 2.2.5.5., and sequenced as described in section 2.2.6..

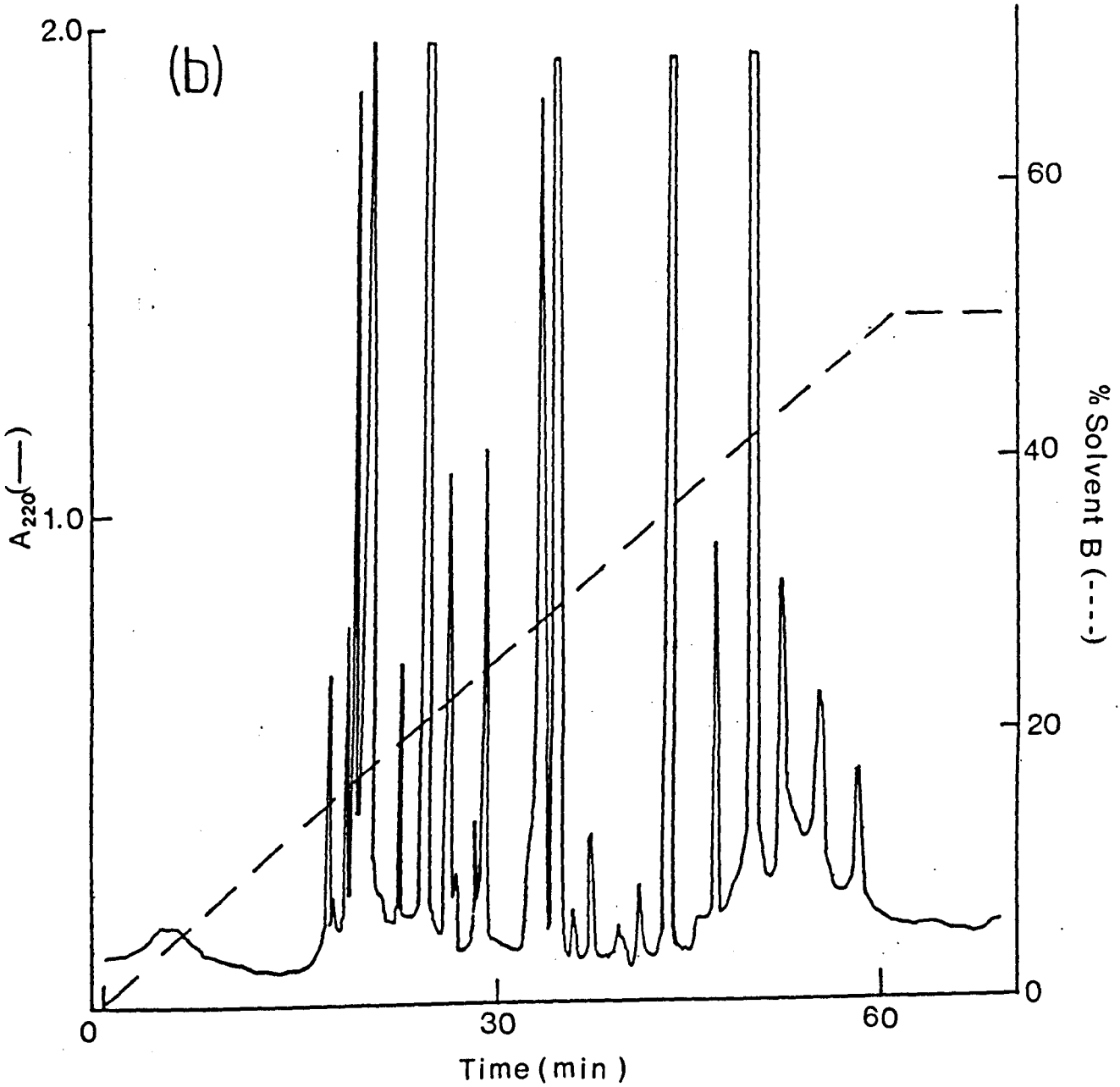


Fig 5.5. (b) PURIFICATION BY H.P.L.C. OF PEPTIDES PRODUCED BY TRYPTIC DIGESTION OF β-LACTOGLOBULIN

The trace shows the digestion profile of unmodified β-lactoglobulin. See the legend to Fig 5.5. (a) for further details.

PEPTIDE 1 L-S-F-N-P-T-Q-L-E-E-Xaa-C_X-H-I

PEPTIDE 2 L-S-F-N-P-T-Xaa-L-E-E-Xaa-C_X-H-I

Fig 5.6. THE SEQUENCE OF TRYPTIC PEPTIDES DERIVED FROM ENZYMATICALLY MODIFIED β-LACTOGLOBULIN [C_X= SULPHONIC ACID DERIVATIVE OF CYSTEINE RESIDUES; Xaa = N-(γ-GLUTAMYL)DANSYLCADAVERINE), OR N-(γ-GLUTAMYL) PUTRESCINE].

The labelled peptides correspond to the sequence Leu_{149} through to Ile_{162} of bovine β-lactoglobulin (Braunitzer et al., 1973). There was a gap in the sequence (Xaa) corresponding to Gln_{159} in peptide 1, and Gln_{155} and Gln_{159} in peptide 2. The [^{14}C]-putrescine modified glutaminyl residue were located at the same positions as the dansylated derivatives within the structure of β-lactoglobulin. However more peptide 2 than peptide 1 was recovered from digests of protein which had been modified with putrescine. This was probably due to: 1) the preferential modification of Gln_{159} by transglutaminase, and 2) the relatively lower degree of steric hindrance caused by the attachment of putrescine compared to that caused by the attachment of the larger dansylcadaverine molecule. It may also be the case that the attachment of two molecules of dansylcadaverine reduced the efficiency of proteolysis of β-lactoglobulin and this also could contribute to lower recoveries of peptide 2.

In consideration of the differences in labelling stoichiometries observed when putrescine and dansylcadaverine are used as substrates, Gln_{155} appears to

be slightly less reactive than Gln_{159}. The further site(s) which became available when incubation was conducted at pH 9.0 was/were not identified. The identity of this site is discussed on the basis of what is known about the relationship between ß-lactoglobulin structure and the specificity of transglutaminase in section 5.3.3..

5.3.2. LOCATION OF MODIFIED SITES WITHIN THE TERTIARY STRUCTURE OF ß-LACTOGLOBULIN

A detailed examination of the X-ray model of bovine ß-lactoglobulin (see Fig 5.7) shows each subunit consists of an antiparallel parallel sheet formed by nine strands wrapped round to form a flattened cone or calyx, with a core consisting of an eight stranded ß-barrel (Papiz et al., 1986). There are two disulphide bonds (Cys_{66} - Cys_{106}) and (Cys_{106} - Cys_{119}) with a free thiol group (Cys_{121}).

Four of the nine glutaminyl residues present in ß-lactoglobulin, [Gln_{13}, Gln_{35}, Gln_{68}, and Gln_{120}] do not have their side-chains significantly exposed to solvent and are therefore unlikely to have substrate properties for transglutaminase when ß-lactoglobulin is in its native state. The side chains of Gln_5 (N-terminal arm), Gln_{59}, and Gln_{115}, appear to be exposed and hence their lack of substrate properties needs to be explained. The electron density about residues Gln_{59}, Gln_{155}, and Gln_{159} is not well defined and hence the local secondary structure associated with these side chains is not known. It is therefore assumed that these sites do have high exposure and considerable local flexibility.

In the presence of DTT one or more of the disulphide bonds are likely to be broken and this would lead to exposure of more sites for transglutaminase reaction. Since the sites of modification at pH 7.6 are confined to Gln_{155} and Gln_{159}, and since

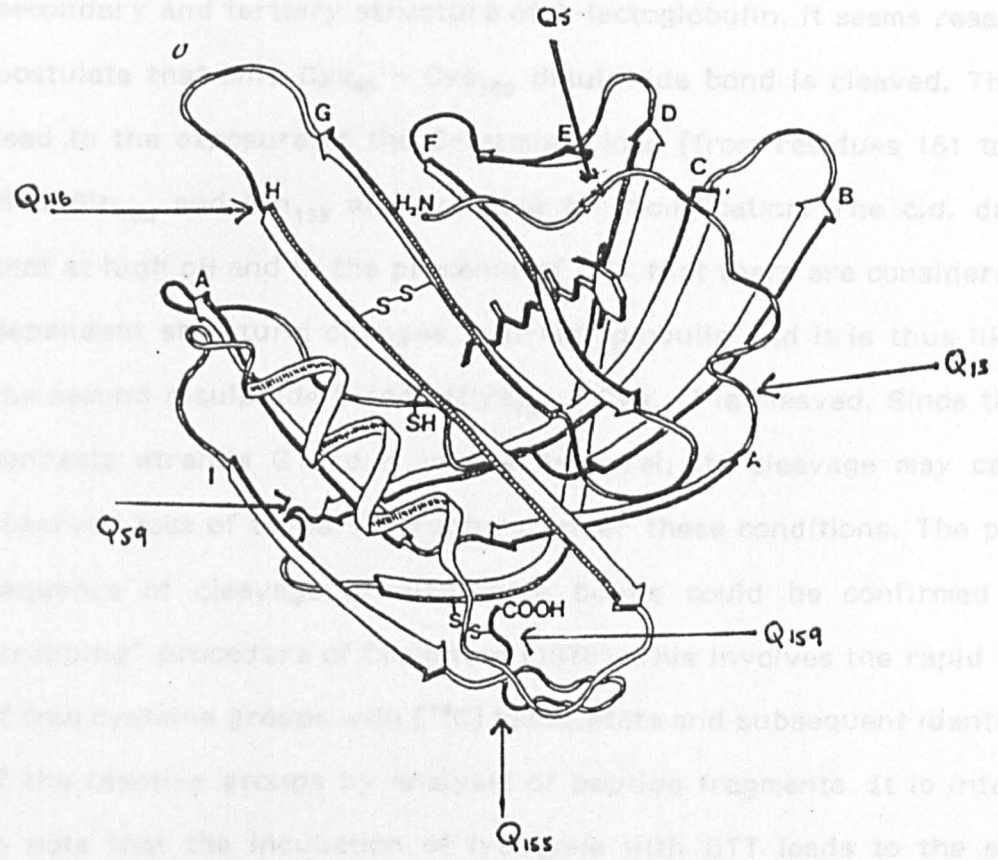


Fig 5.7. THE STRUCTURE OF BOVINE ß-LACTOGLOBULIN

Schematic drawing of the trace of the polypeptide in a subunit of bovine ß-lactoglobulin (Papiz et. al., 1986). Regions of ß-sheet (labelled A to I) are arrow-like bars and α-regions of helix are tubes. At pH 7.6 a significant proportion of the protein is in the monomeric form which is thought to resemble the structure shown above. The position of exposed glutaminyl residues within the structure are indicated by arrows. The disulphide bonds are shown ad S-S structures and the free sulphydryl group is indicated (-SH).

the presence of DTT does not appear to have much effect on the overall secondary and tertiary structure of ß-lactoglobulin, it seems reasonable to postulate that only Cys_{66} - Cys_{160} disulphide bond is cleaved. This would lead to the exposure of the C-terminal loop (from residues 151 to 162) so that Gln_{155} and Gln_{159} are available for modification. The c.d. data show that at high pH and in the presence of DTT that there are considerable time dependent structural changes in ß-lactoglobulin and it is thus likely that the second disulphide bridge (Cys_{106} - Cys_{119}) is cleaved. Since this bond connects strands G and H in the ß-barrel, its cleavage may cause the observed loss of tertiary structure under these conditions. The proposed sequence of cleavage of disulphide bonds could be confirmed by the "trapping" procedure of Creighton (1978). This involves the rapid reaction of free cysteine groups with [^{14}C] iodoacetate and subsequent identification of the reactive groups by analysis of peptide fragments. It is interesting to note that the incubation of lysozyme with DTT leads to the selective cleaveage of one disulphide bond, permitting conjugation of the protein to ubiquitin (Dunten et. al., 1991), a situation analogous to that proposed for ß-lactoglobulin.

5.4.3. THE SPECIFICITY OF GUINEA PIG LIVER TRANSGLUTAMINASE TOWARDS GLUTAMINYL RESIDUES WITHIN ß-LACTOGLOBULIN

The glutamine containing sequences of ß-lactoglobulin are aligned in Table 5.1.. Since there is no evidence for the labelling of any residues apart from Gln_{155} and Gln_{159} at pH 7.6, it is interesting to observe that of the

unlabelled Gln residues all but one of them have positively charged residues on their C-terminal sides only Gln_{115} does not have this motif.

Table 5.1. Alignment of the sequences surrounding glutaminyl residues in bovine β-lactoglobulin

RESIDUE	AMINO ACID SEQUENCE	RESIDUE
1	NH_2-L-I-V-T-**Q**-T-M-K-G-L	9
7	K-G-L-D-I-**Q**-K-V-A-G-T	17
30	S-L-L-D-A-**Q**-S-A-P-L-R	41
54	L-E-I-L-L-**Q**-K-W-E-N-G	64
63	N-G-E-C-A-**Q**-K-K-I-I-A	73
110	S-A-E-P-E-**Q**-S-L-A-C-Q	124
120	Q-S-L-A-C-**Q**-C-L-V-R-T	132
150	S-F-N-P-T-**Q**-L-E-E-Q-C	162
154	T-Q-L-E-E-**Q**-C-H-I-COOH	162

Positively charged residues are in triangles and negatively charged residues are in circles

Positive charge has previously been suggested as a possible inhibitory feature about exposed, but unreactive surface glutamines (Coussons et al. 1991). On this basis it is tentatively predicted that the third labelling site exposed at pH 9.0 is Gln_{115}, which has similarities with a number of other known labelling sites including those reported here, within bovine β-lactoglobulin itself. This site has also been predicted as a possible substrate site on the basis of the cleavage of (Cys_{106} - Cys_{119}) by DTT at high pH (see section 5.3.2.).

From circular dichroism studies the total amount of α-helix within bovine ß-lactoglobulin increases in the presence of DTT. This is indicative of the formation of novel regions of secondary structure following disulphide reduction. Since the labelling studies suggest that the disulphide bridge at the C-terminus of bovine ß-lactoglobulin is the most affected by the presence of reducing agents at pH 7.5, and the c.d. spectra suggest a general retention of the tertiary structure within bovine ß-lactoglobulin it seems possible that in the absence of the restraint imposed on the local secondary structure of the C-terminus by the disulphide bond that this region may refold into a α-helical conformation. A secondary structural prediction of a number of ß-lactoglobulins was carried out by Dr. L Sawyer using the programme PREDICT (see section 9.2.3.). This indicated that the C-terminal region has high α-helical potential which extends from Ser_{150} through to Ile_{162} which contains both of the substrate glutamines determined within this protein. The region from Met_{145} - Ser_{150} is high in ß-turn potential, which if true would be expected to give an "arm" like structure with a flexible "elbow" within the N-terminal region in the partially reduced protein. This is reminiscent of the N-terminal substrate sites within the ß-crystallins, lipocortin I, and porcine phospholipase A2 (see section 1.7.) and the putative C-terminal site within aldolase from rabbit muscle, (see section 4.4.2.).

CHAPTER SIX: AN INVESTIGATION OF THE SUBSTRATE PROPERTIES OF WILD TYPE AND MUTANT PHOSPHOGLYCERATE KINASES (YEAST) TOWARDS TRANSGLUTAMINASE

6.1. INTRODUCTION

Phosphoglycerate kinase (PGK), is a ubiquitous glycolytic enzyme which is one of the best structurally characterized proteins known. The substrate properties of the wild type protein purified from yeast were investigated together with a number of mutant proteins where a single side chains was substituted for another using the technique of site directed mutagenesis. The mutants included; 1) a species where His_{388} was mutated to Gln_{388}, 2) one in which Arg_{168} was mutated to a Lys_{168} residue, and, 3) one in which Arg_{168} residue was mutated to a Met_{168} residue.

METHODS AND MATERIALS

The purified proteins were a kind gift of Dr C.M. Johnson, Department of Chemistry, University of Glasgow. These were produced using plasmids which had been prepared by Dr L. Gilmore, Department of Biochemistry, University of Edinburgh, and Dr H.C. Watson, Department of Biochemistry, University of Bristol. The enzyme activity of PGK was determined at 25°C using the coupled assay conditions described in Appendix II to this thesis. The concentrations of wild type and mutant proteins were determined spectrophotometrically by using an A_{280} for a 1.0 mg/ml solution of 0.49 (Spragg *et al.*, 1976). Circular dichromism studies were conducted on samples as described in section 5.1..

6.2. RESULTS AND DISCUSSION

6.2.1. SOLUTION STRUCTURE OF WILD TYPE AND MUTANT ENZYMES

The near and far c.d. spectra of the wild type and $His_{388} \rightarrow Gln_{388}$ mutant PGK enzymes are shown in Fig 6.1. The far u.v. spectra are identical confirming that they have similar overall secondary structure. Using the CONTIN analysis, (Provencher and Glöckner, 1981) of these spectra over the range 190 - 240 nm, the proportions of amino acids in α-helix and β-sheet structures can be evaluated. The results are 31 % (+/-1 %)α-helix for both enzymes and 33 % (+/- 1 %) and 34 % (+/-1 %) β-sheet for wild type and mutant enzymes respectively. Similar values were obtained for the other mutant proteins tested.

A comparison of the secondary structure predicted from a c.d. analysis of PGK with that calculated from the the X-ray crystal structure (Fig 6.2.) of yeast PGK (Watson et al., 1982), shows that the estimation of α-helix is in good agreement (X-ray structure contains 35 %α-helix). However the percentage β-sheet is considerably greater than that deduced from the X-ray data, (13%). This discrepancy may partly arise from the contribution from β-turns (12 % in the X-ray structure). Since the β-structure makes only a relatively small contribution to the c.d. spectra its estimation is prone to greater error than that of α-helix.

The near u.v. circular dichroism spectra (shown in Fig 6.1a) confirms that the two enzymes possess very similar tertiary structures, although the dichroism of the mutant enzyme is slightly smaller than the wild type. It appears likely that the mutation in the hinge region and the consequent breakage of the putative electrostatic attraction between His_{388} and Glu_{190} causes a small structural perturbation which is detected by neighbouring

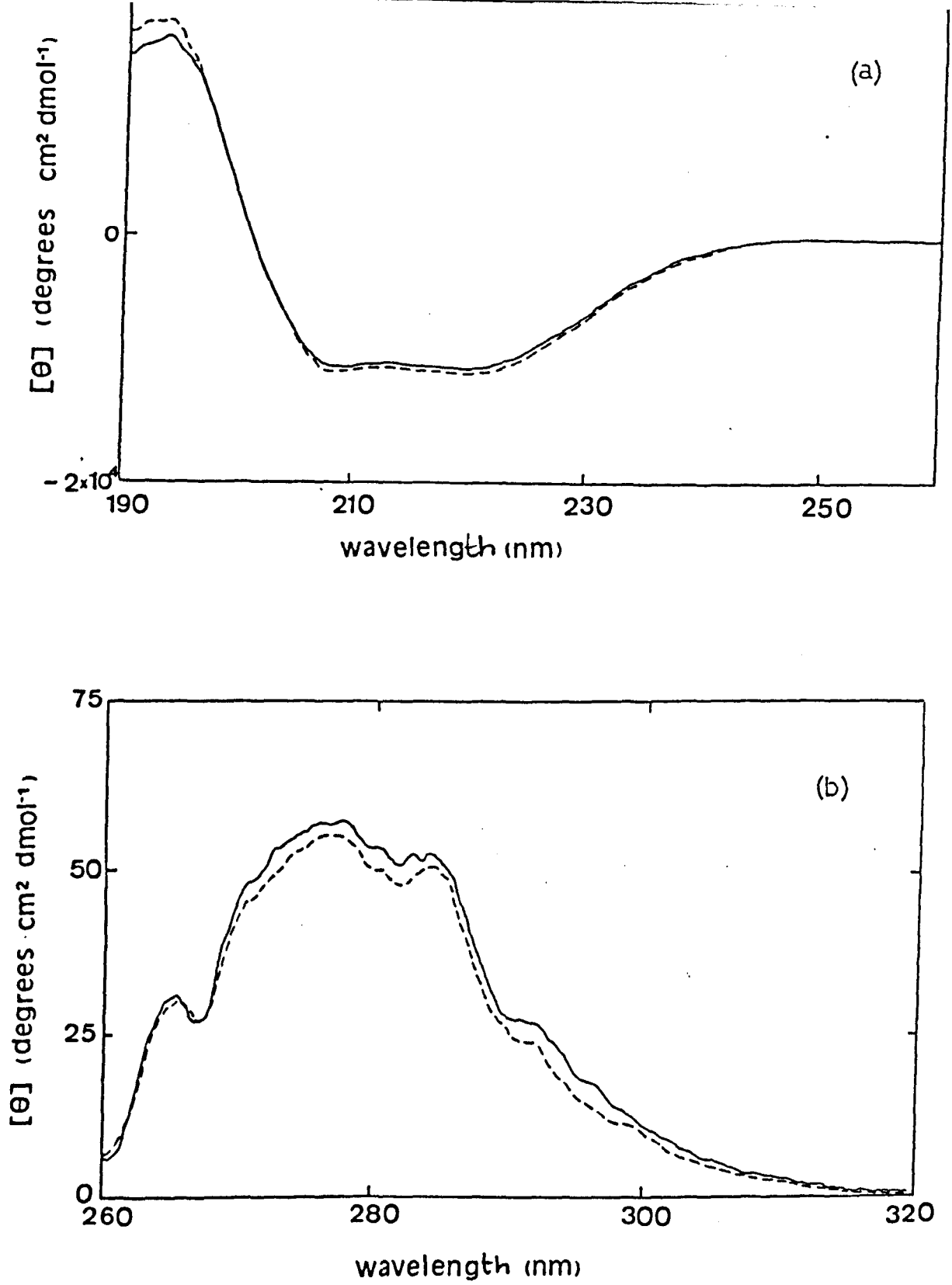


Fig 6.1. C.D. OF YEAST PHOSPHOGLYCERATE KINASE

Spectra were recorded at 20 °C in 25 mM sodium phosphate buffer, pH 7.5.
(a) Far u.v. spectra : (b) near u.v. spectra.
(——) and (- - -) represent wild type and $His_{388} \rightarrow Gln_{388}$ mutant enzymes respectively. The concentrations of enzymes were 0.5 mg/ml and 0.9 mg/ml in (a) and (b) respectively.

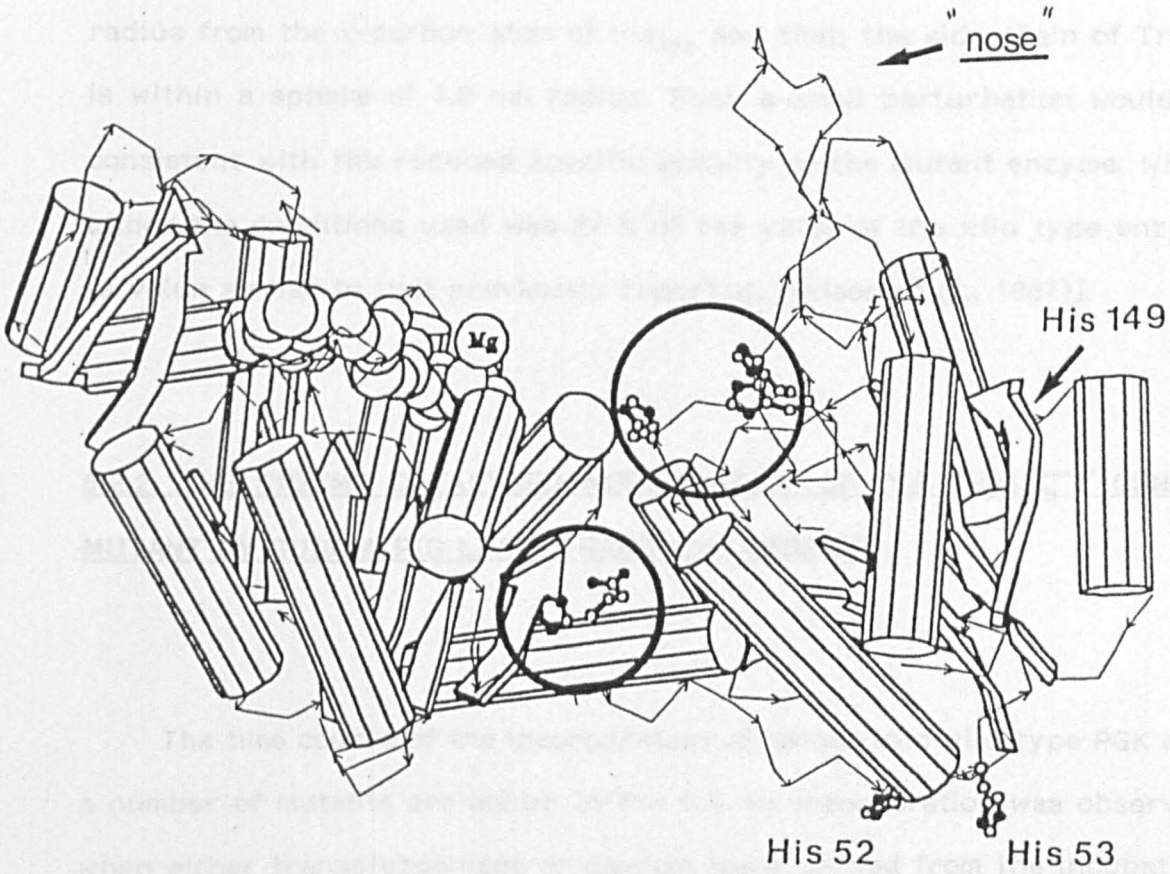


Fig 6.2. THE STRUCTURE OF YEAST PGK

Schematic drawing of the trace of the polypeptide in a subunit of PGK wild type from yeast. Regions of ß-sheet are arrow-like bars and α-helix are tubes. The numbering represents the first and last residues of the amino acid sequence corresponding to each region.The nucleotide substrate is shown in space filling mode. The histidine residues located in the basic patch region are in the upper circle. His_{388} and the residue to which it is H-bonded (Glu_{190}) are shown in the lower circle. In the mutant protein tested His_{388} has been converted to a glutaminyl residue. The surface loop which corresponds to the nose region is indicated by an arrow. The diagram is from Graham et al., 1991).

side-chains. Inspection of X-ray structure of PGK (Watson et al., 1982) shows that, side chains of Tyr_{193} and Phe_{194} are within a sphere of 0.75 nm radius from the α-carbon atom of His_{388} and that, the side chain of Trp_{333} is within a sphere of 1.0 nm radius. Such a small perturbation would be consistent with the reduced specific activity of the mutant enzyme, which under the conditions used was 21 % of the value of the wild type enzyme [a value similar to that previously reported, (Wilson et al., 1987)].

6.2.2. THE ENZYME CATALYSED MODIFICATION OF PGK ($HIS_{388} \longrightarrow GLN_{388}$) MUTANT BY GUINEA PIG LIVER TRANSGLUTAMINASE

The time course of the incorporation of amines into wild type PGK and a number of mutants are shown in Fig 6.3. No incorporation was observed when either transglutaminase or calcium were omitted from the incubation mixture. The reaction was not dependent on the presence of DTT and this is reasonable given the lack of disulphide bridges in all of the PGK variants tested. The wild type enzyme proved to be a poor substrate for transglutaminase. Less than 0.1 mol putrescine/mol incorporation was observed after 6 hr rising to 0.2 mol putrescine/mol after 24 hr incubation. It is possible that this incorporation was due to partial unfolding of the enzyme, since there was a corresponding slow decline in the specific activity of the enzyme under these conditions, both in the presence and in the absence of transglutaminase. Over this period there was no significant change in the near u.v. spectra of the wild type enzyme (see Fig 6.1b).

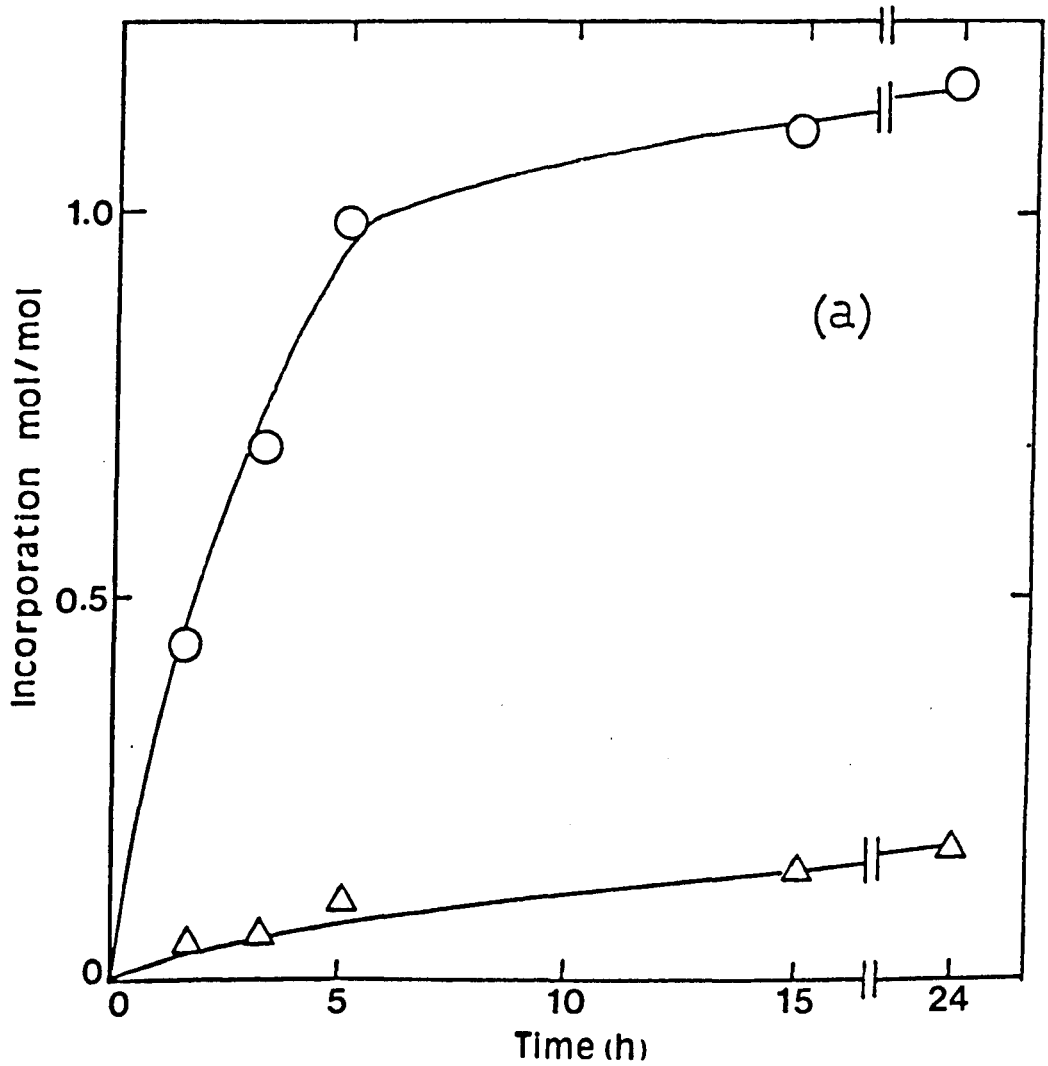


Fig 6.3. THE MODIFICATION OF PGK BY TRANSGLUTAMINASE

Transglutaminase was used to incorporate [^{14}C] putrescine into wild type PGK and a number of mutant forms. Reaction conditions are described in the methods section.

(△) and (○) represent data for the wild type and the $His_{388} \longrightarrow Gln_{388}$ mutant enzymes respectively. In control experiments where transglutaminase was omitted from the incubation mixture there was 0.02 mol amine/mol PGK incorporated after 24 hr. No significant incorporation was observed in the other mutant forms of PGK.

When the $His_{388} \longrightarrow Gln_{388}$ mutant enzyme is used as substrate for the transglutaminase catalysed reaction there is rapid incorporation of [^{14}C] putrescine to a stoichiometry of 1.1 mol putrescine/mol protein after 6 hr and to about 1.3 mol putrescine/mol protein after 24 hr. Like the wild type there is a small decline in specific activity over this time period but this occurs both in the presence and absence of transglutaminase catalysed labelling and therefore is not a consequence of the labelling reaction (see Fig 6.4). No significant incorporation was noted in the other mutant PGK enzymes tested.

The incorporation of monodansylcadaverine into PGK was recorded after 20 hr reaction time after removal of excess amine by gel filtration (see section [2.2.2.3.(b)]. The extent of incorporation into the wild type and His_{388} mutant enzyme were 0.3 and 1.26 mol dansylcadaverine/mol protein respectively, ie. very similar to the stoichiometries recorded with putrescine. The other mutants tested showed low degrees of incorporation comparable with the wild type. Transglutaminase catalysed dansylation of the PGK $His_{388} \longrightarrow Gln_{388}$ mutant did not appear to reduce the specific activities of either protein to an extent greater than that observed in the controls.

SDS-PAGE of unmodified and modified (by either dansylcadaverine or putrescine) proteins showed that there was no detectable breakdown of polypeptide chain or formation of higher aggregates during incubation for 20 hr, ie. that there were no detectable proteinases in the preparations and that cross-linking reactions were not favoured under the conditions used.

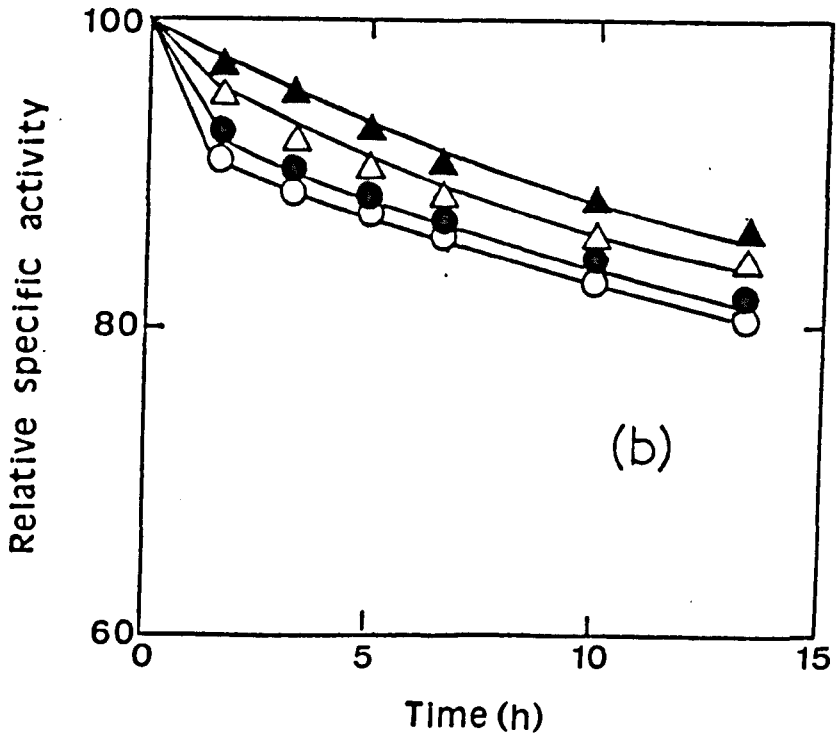


Fig 6.4. THE EFFECT OF PGK MODIFICATION ON PGK SPECIFIC ACTIVITY

(O) and (Δ) represent the data for the wild type and $His_{388} \longrightarrow Gln_{388}$ mutant enzymes respectively. The corresponding closed symbols represent control mixture from which transglutaminase was omitted. The initial specific activities employed were 436 and 93 μmol/min/mg for wild type and mutant enzymes respectively.

6.2.3. LOCATION OF THE SITE OF MODIFICATION FOR TRANSGLUTAMINASE WITHIN THE PGK (HIS $_{388} \rightarrow$ GLN $_{388}$) MUTANT STRUCTURE

6.2.3.1. LOCATION OF THE MODIFIED SITE WITHIN THE PRIMARY STRUCTURE

Dansylated PGK (1.0 mg/ml) was digested in 1% ammonium bicarbonate with 1:40 w/w TPCK treated bovine pancreatic trypsin for 4.0 h and a modified peptide was identified and purified by HPLC using a gradient of isopropanol in water (TEA 1% v/v) as described in section 2.2.5.5.. Labelled peptide was distinguished from free dansylcadaverine, which was observed to be eluted at 20 % solvent B by the use of a dansylcadaverine standard. The HPLC profiles are shown in Fig 6.5. A comparison of the maps generated for the modified His_{388} mutant enzyme shows that there is a distinct peak (labelled "P") containing the bulk of the fluorescence which eluted at a position corresponding to 45 % solvent B. This peak was not present in the tryptic digest of the unmodified protein and this suggested that this peptide represented the major site of labelling. A small peak was observed at 20 % B probably reflecting a small amount of dansylcadaverine which is non-specifically bound even after prolonged dialysis. It should be noted that the HPLC profile of the digest of the unmodified enzyme displayed less well resolved peaks in the region corresponding to 35 - 40% solvent B than did the unmodified protein. This suggests that the presence of the large hydrophobic dansylcadaverine moiety may inhibit the action of trypsin on neighbouring peptide bonds (this phenomenon has been discussed in section 5.3.1.).

The material corresponding to peak "P" was further purified by HPLC on an Aquabore butyl column using an increasing gradient of acetonitrile

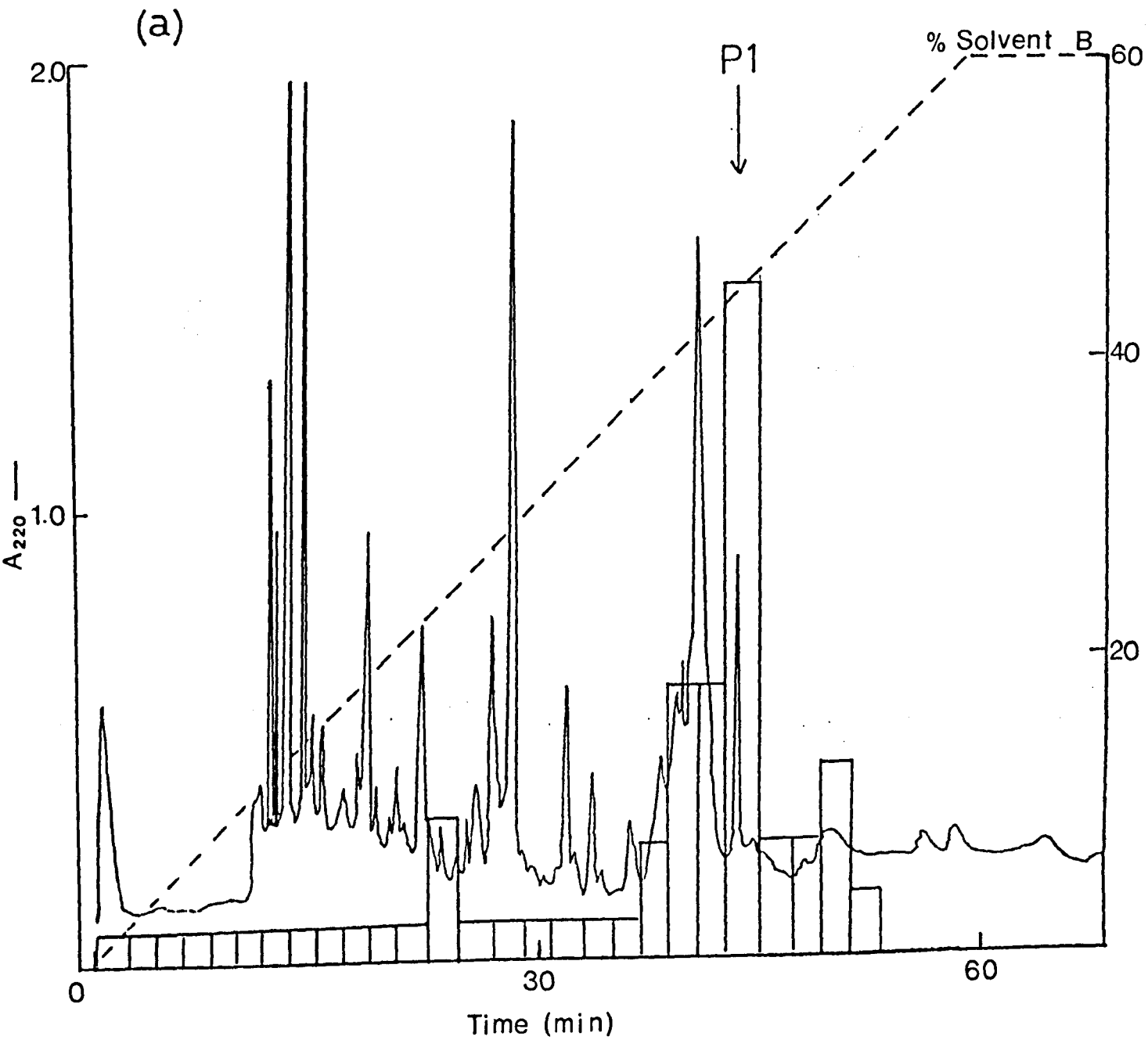


Fig 6.5. (a) PURIFICATION BY H.P.L.C. OF PEPTIDES PRODUCED BY TRYPTIC DIGESTION OF PGK

The modification of the $His_{388} \longrightarrow Gln_{388}$ mutant of PGK was performed as described in the experimental section. Both traces (a) and (b) refer to separations of peptides on a Vydac C-18 column for dansylcadaverine modified and unmodified protein respectively. In each case approximately 0.2 mg of digestedprotein was applied. Solvent A consisted of aqueous 0.4% triethylamine, adjusted to pH 2.5 with orthophosphoric acid. Solvent B was solvent A containing 60% (v/v) propan-2-ol. In (a) the rectangles show the specific fluorescence of the fractions (excitation and emmision wavelengths 320 and 520 nm respectively).

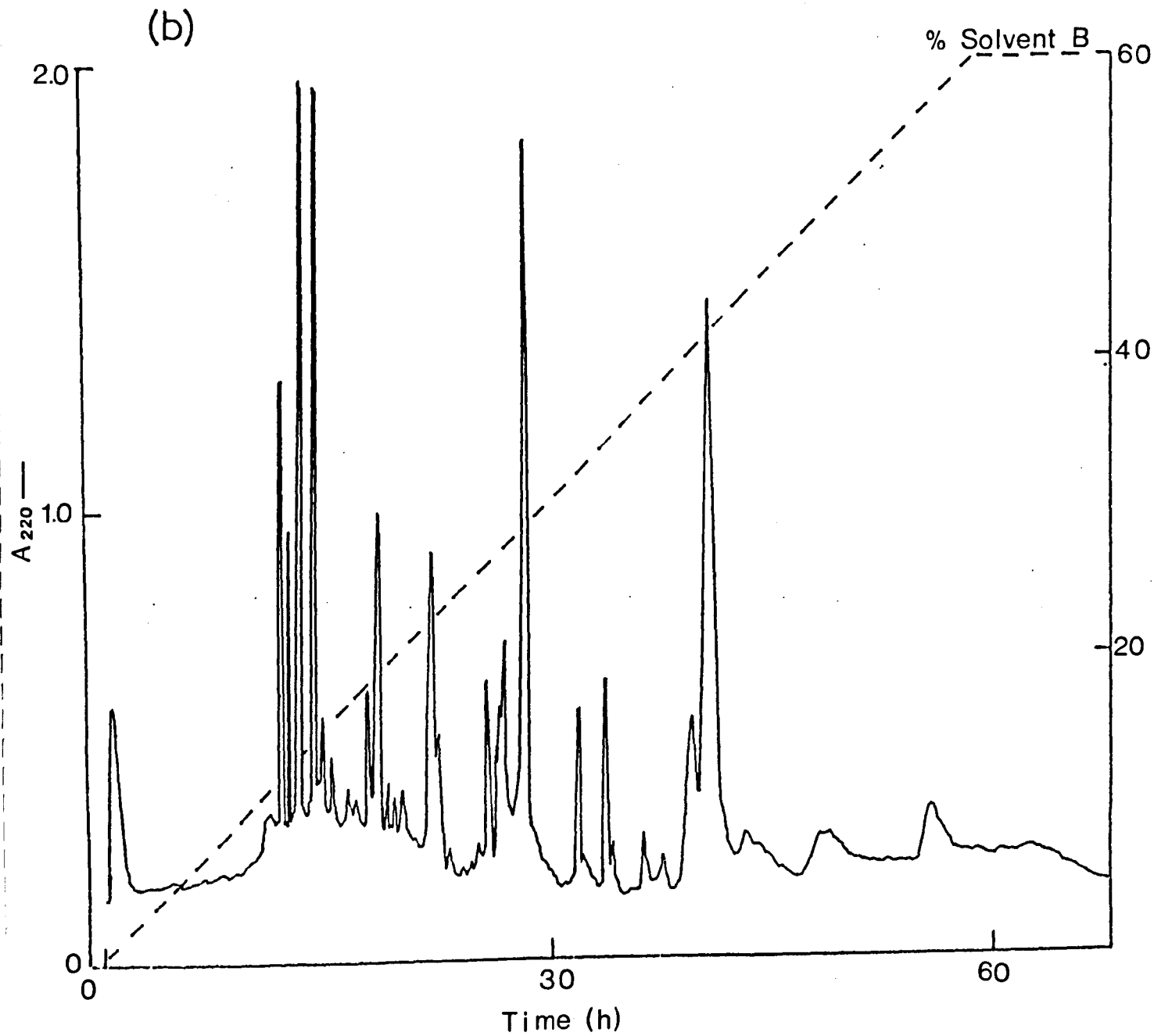


Fig 6.5. (b) PURIFICATION BY H.P.L.C. OF PEPTIDES PRODUCED BY TRYPTIC DIGESTION OF PGK

The HPLC profile of the tryptic peptides of unmodified PGK is shown above. See Fig 6.5 (a) for further details.

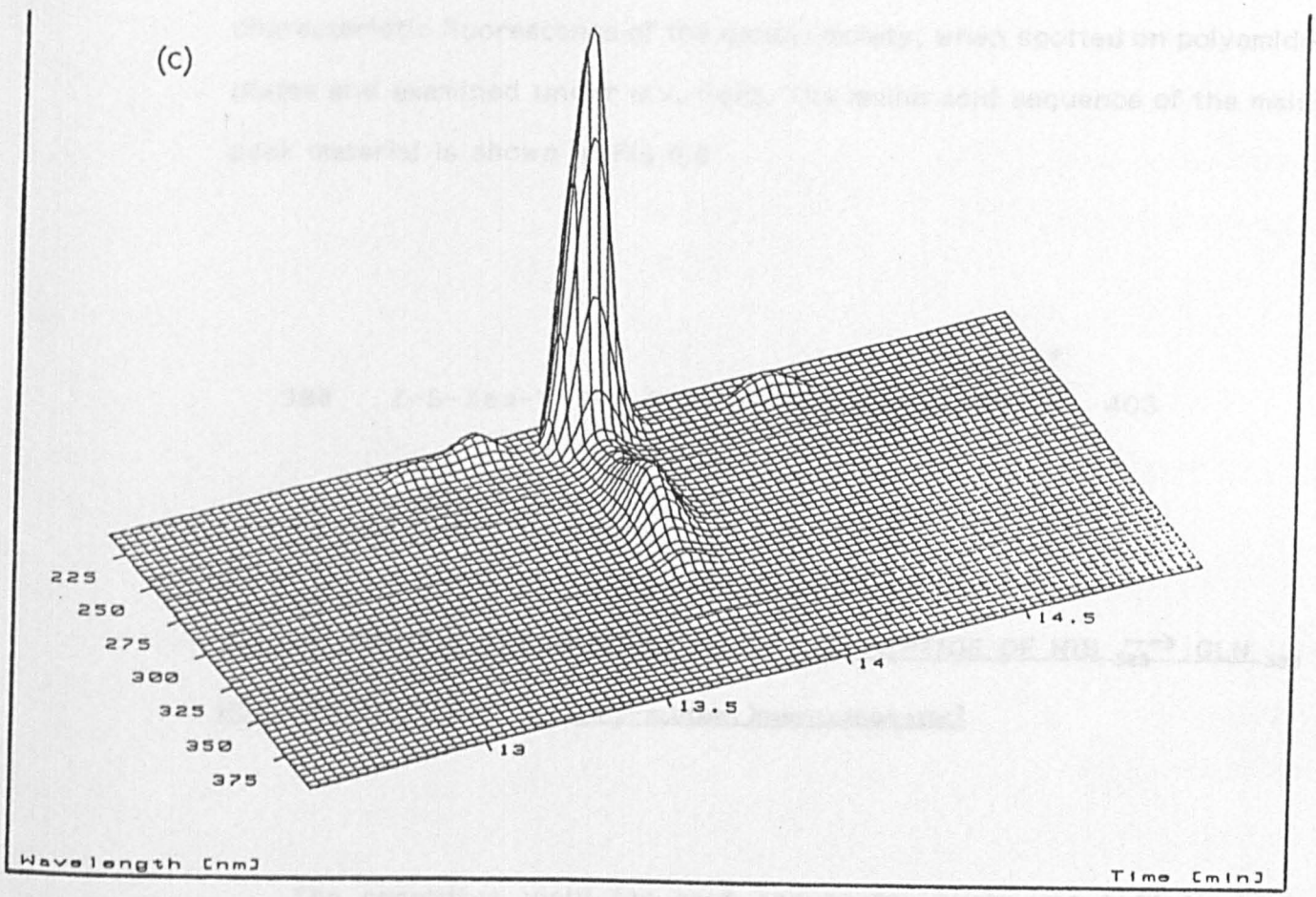


Fig 6.5. (c) PURIFICATION BY H.P.L.C. OF PEPTIDES PRODUCED BY TRYPTIC DIGESTION OF PGK

The major fluorescent peak seen in Fig 6.5 (a) was rechromatographed on an aquabore column (see text) and the output from the diode array detector is shown above. The axis perpendicular to the hatched plane represents absorbance. The peptide shows significant absorbance at 330 nm which is typical of that of the dansyl group of dansylcadaverine.

in water (TFA 1% v/v). The main peak, corresponding to 32 % solvent B accounted for 90 % of the applied material, as judged by its absorbance at 220 nm. Diode array detection indicated that the pure peptide had an absorbance maxima at 330 nm (see Fig 6.5c) and this material showed the characteristic fluorescence of the dansyl moiety, when spotted on polyamide plates and examined under u.v. light. The amino acid sequence of the main peak material is shown in Fig 6.6

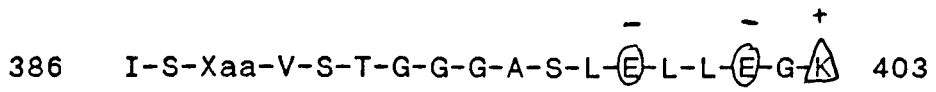


Fig 6.6. TRYPTIC FRAGMENT P1: LABELLED PEPTIDE OF HIS $\xrightarrow{388}$ GLN $_{388}$ PGK (YEAST) [-X_{AA}- = N-ε(γ-GLUTAMYL)DANSYLCADAVERINE]

The repetitive yield for each sequenator cycle was > 85 %. This sequence corresponds to the unique sequence of amino acids Ile_{386} through to Lys_{403} in the published amino acid sequence of yeast PGK determined from the protein (Perkins et al., 1983), and from the gene (Hitzeman et al., 1982). The unknown amino acid (Xaa) in the sequence corresponds to Gln_{388} in the mutant enzyme, and this produced a charactaristic gap in the sequence (see also section 5.3.1.). The PTH derivatives produced during sequencing were freeze dried and a portion of each applied to a polyamide plate; the sample which corresponded to Gln_{388} within the sequence showed

the characteristic pale green fluorescence of the dansyl group when excited with near u.v. radiation.

As a final check that the dansylcadaverine moiety had been incorporated into the peptide in the mutant enzyme, the material P1 was analyzed by fast atom bombardment (FAB) mass spectrometry. This showed that the parent ion had a Mr of 2062 Da as predicted from the sequence of amino acids in the peptide containing the monodansylcadaverine derivative of glutamine, see Fig 6.7. The F.A.B. mass spectrometry experiments were done by Mr T. Gibson (Xenova Ltd.).

6.2.3.2. LOCATION OF GLN_{388} WITHIN THE TERTIARY STRUCTURE OF PGK

The c.d. analysis of the $His_{388} \longrightarrow Gln_{388}$ mutant (Fig 6.1 a,b.) showed that it was very similar in overall tertiary structure to the wild type for which the X-ray structure is available, (Watson et al., 1982). The mutation is considered to lie within the so called "hinge" region. This part of the protein lies within a connecting region between the two globular domains of PGK and may affect the catalytic movement of these domains relative to one another. The c.d. spectra confirm that the mutation may cause a small structural perturbation in the structure of PGK. The mutation is reported to have an overall "loosening" effect on the tertiary structure of the mutant protein relative to the wild type (Graham et al., (1991). It appears possible that as a result of the breaking of the His_{388} --- Glu_{190} salt bridge, a stretch of the polypeptide chain may loop out increasing access of Gln_{388} to the active site of transglutaminase.

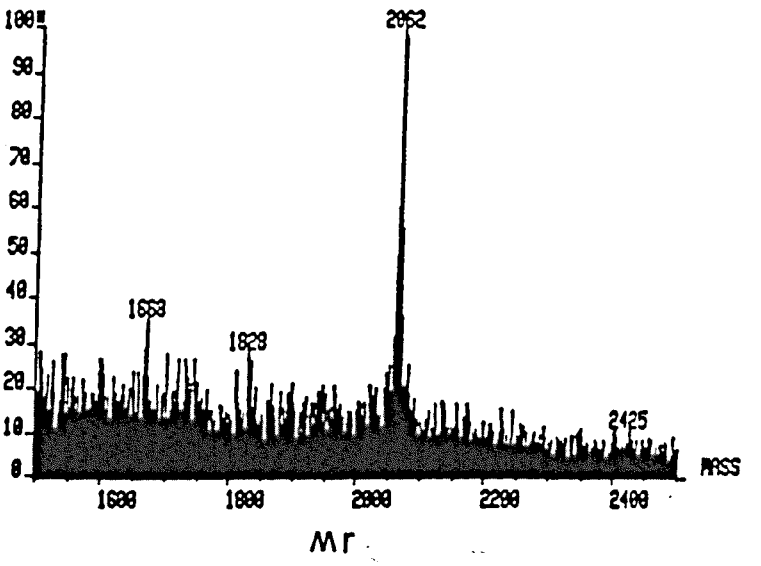


Fig 6.7. FAB MASS SPECTROMETRY OF PEPTIDE P1

The purified fluorescent peptide (see Fig 6.5.) was subjected to fast atom bombardment mass spectrometry. The trace is shown above and indicates that the modified peptide has a molecular weight of 2052 D. This is the same molecular weight calculated for the dansylcadaverine derivative of the peptide sequence determined from the Edman degradation of peptide P.

The modified site in PGK has some sequence (Korn, 1982) similarity with the region which surrounds the reactive Gln_{41} (Takashi, 1988) of rabbit muscle actin. This region forms an externally located fourteen residue loop structure in the actin monomer (Kabsch et al., 1990), which has also been shown to have substrate properties (Takashi, 1988). It is tempting to speculate on the basis of its similar substrate properties that the mutation of His_{388} to Gln_{388} causes the formation of a loop structure which resembles that observed in the X-ray structure of actin. The amino acid sequences surrounding these modification sites are compared in Fig 6.8.

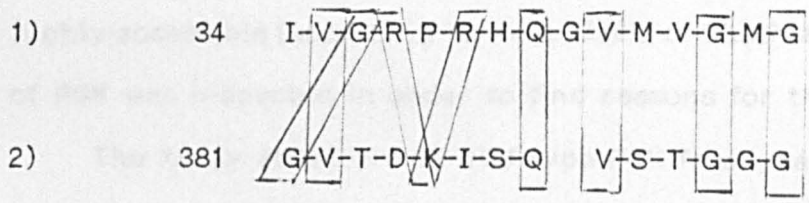


Fig 6.8. A COMPARISON OF LABELLING SITES IN: 1) ACTIN, 2) PGK (GLN_{388}) MUTANT [a gap has been artificially introduced into the PGK sequence to increase the quality of alignment], residues in eqivalent positions in both sequences are boxed

6.3. THE SPECIFICITY OF TRANSGLUTAMINASE TOWARDS PGK

Only minor levels of amine were incorporated by transglutaminase into wild type PGK and the $Arg_{168} \rightarrow Met_{168}$, and $Arg_{168} \rightarrow Lys_{168}$ mutants. These proteins were also judged by circular dichroism to have very similar overall structures to the $His_{388} \rightarrow Gln_{388}$ mutant, but clearly lacked suitable sites for modification by transglutaminase. The unreactive glutaminyl residues thus probably have similar local conformations in each mutant form of PGK as well as identical surrounding amino acid sequences. Since a number of glutaminyl residues within the wild type protein are known to be highly accessible [according to the criteria of Chothia, (1975)] the structure of PGK was inspected in order to find reasons for this lack of reactivity.

The X-ray structure of wild type PGK from yeast is shown in Fig 6.2. The "nose" region is indicated by an arrow and is a flexible glutamine-rich surface exposed loop of approximately 20 amino acids in length. Despite the high accessibility of the glutaminyl residues within this region, none of them appear to be reactive towards transglutaminase. By comparison with the reactive loop within the structure of actin from rabbit muscle, (see section 1.7.8.3.) this portion of the molecule should be of adequate dimensions for modification by transglutaminase. The lack of labelling in this region, may therefore be due to discouraging factors within the linear sequences which surround the glutaminyl residues located within this region. The sequence of this region (Perkins et al., 1983) is shown in Fig 6.9.

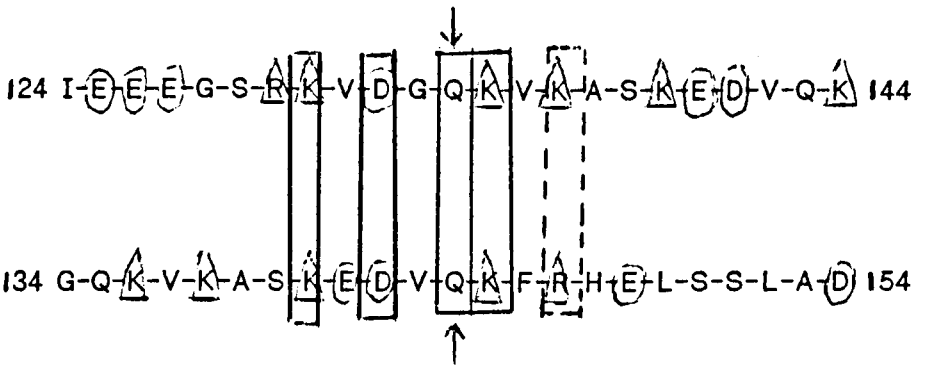


Fig 6.9. SEQUENCES SURROUNDING UNREACTIVE GLN$_{135}$ AND GLN$_{145}$ SITES IN PGK positively charged residues are in triangles and negatively charged residues are in circles. Residues in identical positions in both sequences are boxed

The presence of positively charged side-chains about the glutaminyl residues (particularly on the C-terminal side) may discourage the transglutaminase catalysed reaction. A similar lack of modification of glutamine residues which have positively charged residues adjacent to them in the local peptide sequence has also been noted in bovine ß-lactoglobulin, (see section 5.3.3.), and in a number of other protein substrates (see section 1.7.8.4.).

CHAPTER SEVEN: AN INVESTIGATION INTO THE SUBSTRATE PROPERTIES OF BOVINE ß-CASEIN TOWARDS TRANSGLUTAMINASE

7.1. INTRODUCTION

The substrate properties of ß-casein have been investigated previously by Yan and Wold (1984). The substrate sites (see section 1.7.7.3.) were determined in material in which all the lysine side chains of bovine ß-casein were modified with succinic anhydride in order to prevent inter- and intra-molecular cross-linking reactions. This treatment may have conformationally altered the substrate to expose more sites to the enzyme than are available in its non-succinylated form. The major modification sites for factor XIIIa within non-succinylated ß-casein have been determined and found to be Gln_{167}, but the substrate sites for the guinea pig liver enzyme are not known for the native protein. The substrate properties of non-succinylated bovine ß-casein are therefore considered in sections 7.2. - 7.4. and the specificity of guinea pig liver transglutaminase towards this substrate is considered in section 7.5.. The ß-casein used for the following experiments was purchased from Sigma Chemical Co.

7.2. RESULTS AND DISCUSSION

7.2.1. SOLUTION STRUCTURE OF ß-CASEIN (BOVINE)

The far u.v. spectra of ß-casein both in the presence and absence of Ca^{2+} (aq) at pH 7.6 (ie. that used for the enzyme catalysed modification of this protein) are shown in Fig 7.1.. A CONTIN (Provencher and Glöckner, 1981) analysis of the spectra recorded under these conditions

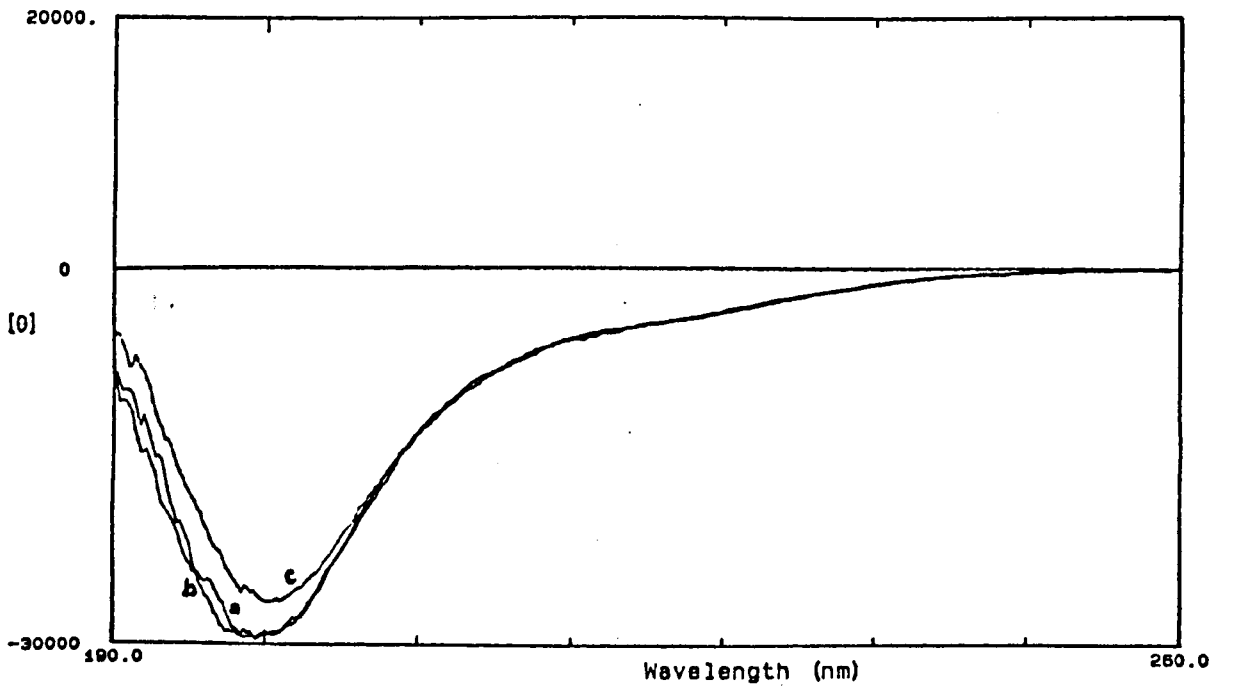


Fig 7.1. FAR U.V. C.D. SPECTRA OF BOVINE β-CASEIN

Spectra were recorded under the standard conditions described in the text ie. at 20 °C in 50 mM tris-HCl pH 7.5, and in 25 mM sodium phosphate buffer, pH 7.5. The spectra were identical to (a) in each case. The concentration of casein was 1.0 mg/ml.The spectra represent the average of three blank corrected spectra in each case. A CONTIN (Provencher and Glockner, 1981) analysis was done on this data in order to predict the content of secondary structure in the sample under these conditions.

(a) In the absence of calcium.

(b) In the presence of 5 mM Ca^{2+}

(c) The same as (b) following an overnight incubation.

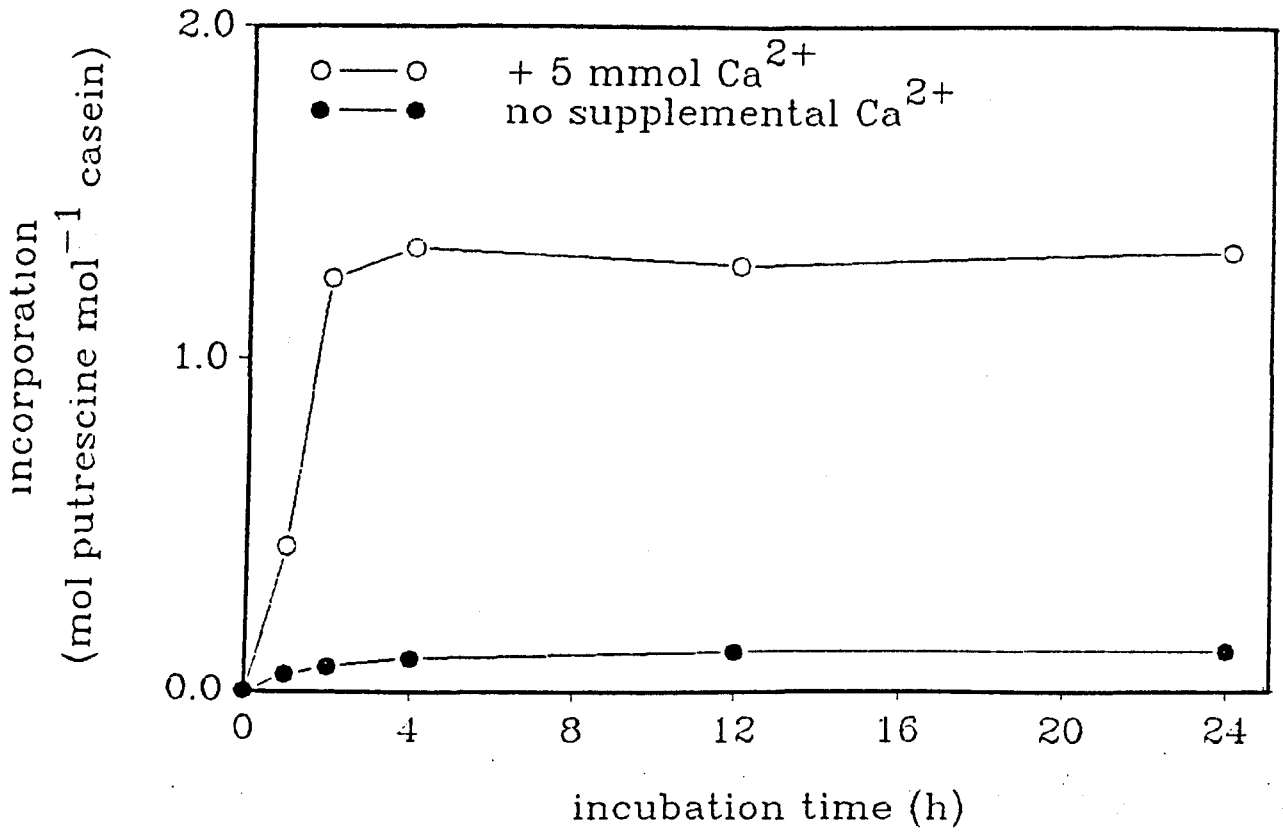


Fig 7.2. TIME COURSE OF INCORPORATION OF PUTRESCINE INTO β-CASEIN BY TRANSGLUTAMINASE

The concentration of β-casein was 2 mg/ml and the incubations were conducted at 25°C using the standard reaction conditions described in section 2.2.3.2.. The buffer used was triethanolamine (50 mM) pH 7.5. Further details are given in the text.

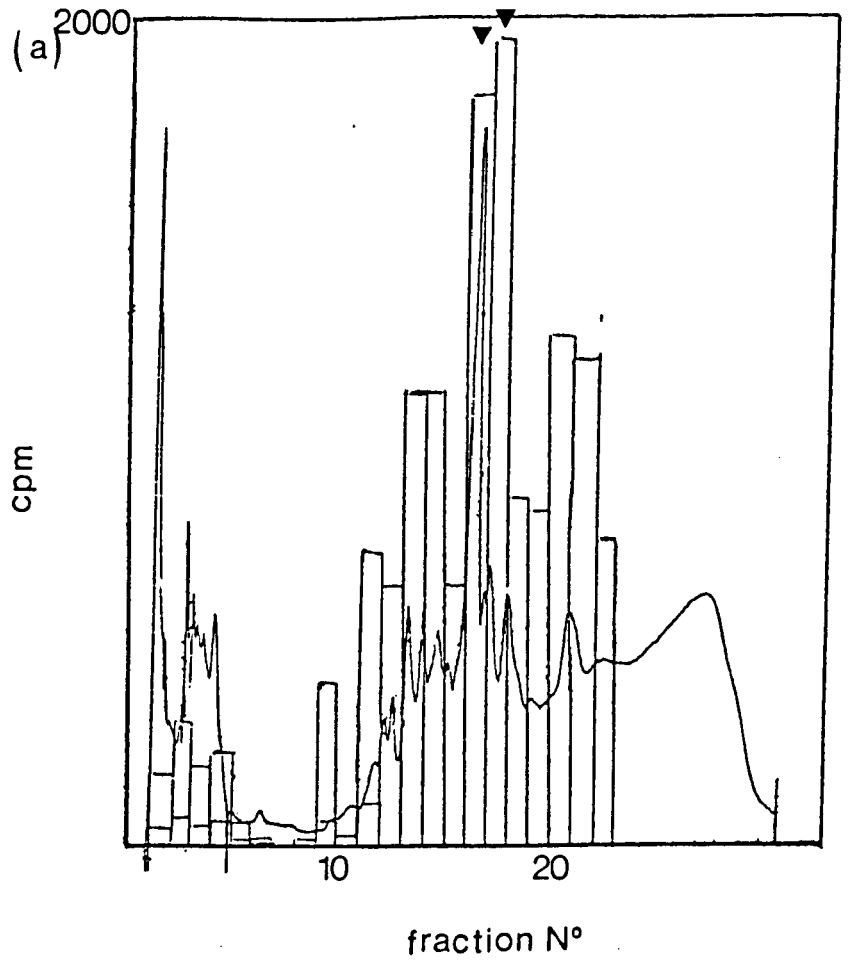


Fig 7.3. HPLC PROFILE OF THE SEPARATION OF [^{14}C] PUTRESCINE LABELLED PEPTIDES OF BOVINE β-CASEIN

(a) β-casein which had been modified to a stoichiometry of 1.5 mol/mol with putrescine using transglutaminase was carboxymethylated (see section 2.2.5.2.) and digested with TPCK treated trypsin as described in section 2.2.5.3.. The peptide fragments were separated over 40 min using a PEP RP1 (Dupont) reverse phase HPLC column using a 5.0 - 65 % gradient of acetonitrile in water, with 0.1 % v/v TFA in each phase. The flow rate was 0.5 ml/min. The sample was washed for 5.0 min with solvent A prior to application of the gradient. This removed unbound putrescine. Two 100 ul aliquots of 1% ammonium bicarbonate were then applied and the sample was washed again for 10 min, this procedure improved the peptide yield by up to 50 % compared to separations when this step was omitted. The yield of radioactivity recovered from the peptides was >90% of that loaded in the original sample. The rectangles indicate the amount of radioactivity (background corrected) which was present in each 1.0 ml aliquot of the eluate. A similar digestion of unmodified β-casein gave a very similar trace since hence it appears that the covalent attachment of putrescine had little effect on the binding behaviour of the modified peptide on this column. The major radioactive peptides(▼) were collected and rechromatographed [see Fig 7.3 (b)]

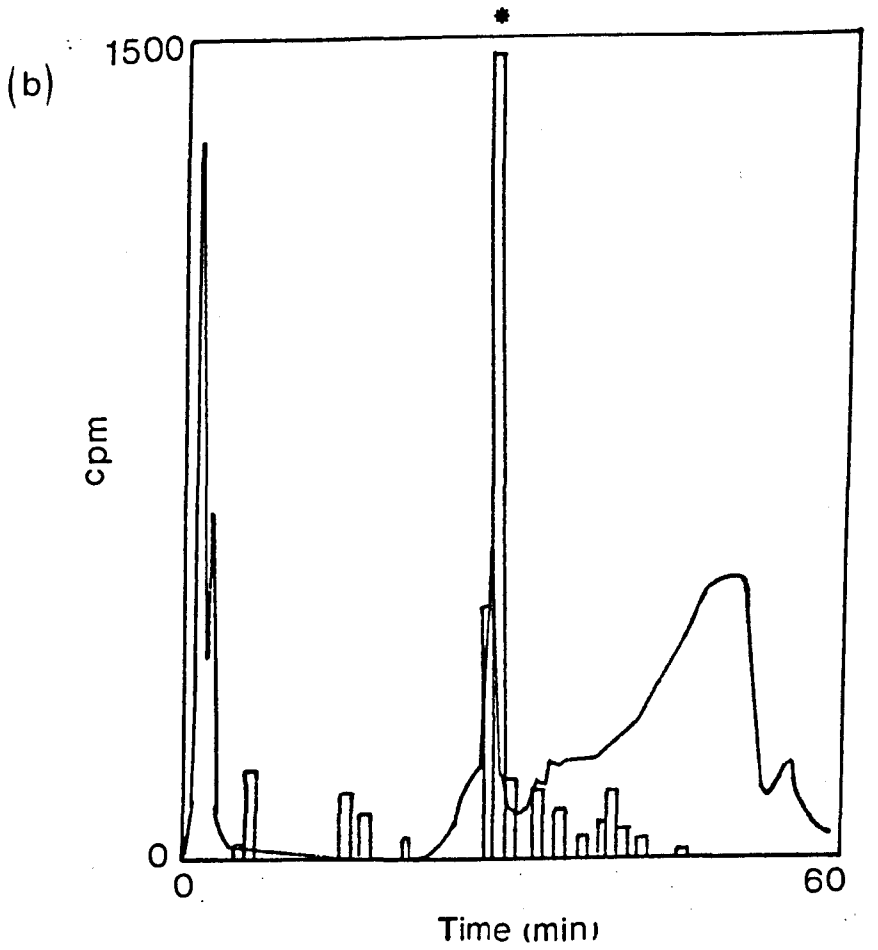


Fig 7.3. HPLC PROFILE OF THE SEPARATION OF [^{14}C] PUTRESCINE LABELLED PEPTIDES OF BOVINE β-CASEIN

(b) Rechromatography of the major radioactive peptide (*) which was recovered from separation (a). The gradient used was the same as used in separation (a). Over 75 % of the radioactivity of the sample applied to the column was recovered in the peptide peak which appeared at the same position in the chromatogram as in Fig (a). This was directly sequenced and fractions collected at each cycle of the degradation were tested for radioactivity.

indicated the following secondary structural content : [α-helix 24 % +/-2 %, ß-sheet 25 % +/- 2 %, and other 51 % +/- 3 %]. These figures may be compared with the secondary structural prediction of ß-casein, determined by Holt and Sawyer (1988) which suggested a composition of 20 % α-helix, 5 - 7 % ß-sheet, 4 % ß-turn and approximately 69 % other structure. Unlike the other substrates tested for substrate properties, bovine ß-casein is thought to exist as multimeric complexes in solutions containing calcium, and this will act as a constraint to the flexibility of the polypeptide chain compared to samples which are incubated in the absence of calcium (Sawyer et al., 1988).

7.2.2. THE ENZYME CATALYSED MODIFICATION OF BOVINE ß-CASEIN BY TRANSGLUTAMINASE

The time-course of putrescine incorporation into ß-casein is shown in Fig 7.2. The reaction is fully dependent on the presence of both calcium and transglutaminase, since there is little or no radioactivity incorporated into protein in control incubations which exclude either of these components. The stoichiometry of labelling was consistently between 1.0 - 2.0 mol putrescine/mol ß-casein depending on the concentration of ß-casein used, in the range 1 - 5 mg/ml, lower stoichiometries of modification being observed at higher protein concentrations. The stoichiometry was not increased by further addition of transglutaminase. This indicated that all of the available sites were fully labelled. The degree of cross-linking of ß-casein into higher polymeric aggregates was not observed to be a predominant reaction (the band at 24 kDa corresponding to ß-casein was

retained during reaction, although a proportion of higher Mr material was observed on SDS gels after 20 hr incubation. This was judged to be approximately 10 % of the starting material and appeared to predominantly consist of dimeric material. Hence it appears that the high concentration of putrescine (5 mM) effectively competed against the exposed lysine residues in ß-casein for the active site of transglutaminase.

The stoichiometry of incorporation of label into ß-casein was lower than that observed by Yan and Wold (1984) who reported 5 mol amine/mol protein incorporation into amidated ß-casein and 8 mol amine/mol protein into succinylated ß-casein. Brenner and Wold (1978), have reported a stoichiometry of incorporation of 2.0 mol methylamine/mol protein with non-succinylated ß-casein using guinea pig liver transglutaminase.

7.3. LOCATION OF A MODIFIED SITE WITHIN THE PRIMARY STRUCTURE OF BOVINE ß-CASEIN

Peptide mapping by HPLC (Fig 7.3.) indicated that there was one major site of labelling and probably 2 or 3 other peptides which were modified to between 25 to 75 % of the level of the main peak. The most highly radioactively labelled peptide fragment from a tryptic digest of ß-casein was retained. The sample was reduced in volume by 90 % by use of a rotary evaporator (Gyrovap) at 37°C. The sample was then rechromatographed by HPLC on a PEP RP 1 column after being washed on the column with 3 x 100 µl aliquots of ammonium bicarbonate which gave a recovery of > 90 % of the initial radioactivity. The material which eluted at approximately 40 % solvent B was sequenced and the resulting 29 residue peptide is shown in Fig 7.4.

Fractions corresponding to each cycle of the sequencer were collected, mixed with 4 ml of scintillation fluid and the radioactivity associated with these samples was determined. A peak of radioactivity was observed in cycle 11 of the sequence determination of the purified peptide which corresponded to Gln_{79}, but no radioactivity was detected in cycles 4 or 21 which corresponded to the positions of the other glutaminyl residues (Gln_{72}, and Gln_{89}) within the sequence. The pattern of reactivity of glutaminyl residues within this region of ß-casein is identical to that determined by Yan and Wold, (1984) using the succinylated derivative of ß-casein.

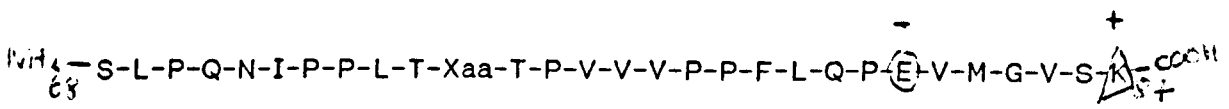


Fig 7.4. TRYPTIC FRAGMENT P1: LABELLED SITE WITHIN NATIVE BOVINE ß-CASEIN [Xaa = N-ε(γ-glutamyl)putrescine] positively charged residues are in triangles and negatively charged residues are in circles

This peptide corresponds to the region Ser_{68} through to Lys_{87} within the sequence of bovine ß-casein (Yoshimura et al., 1986). The peptide was the product of a pseudo-tryptic cut site between Asn_{67} and Ser_{68}. The sequence lacks charged residues in the vicinity of the labelling site and is highly hydrophobic and proline rich. It is also characterised by the presence of neighbouring threonine (hydroxylated) residues about the

presence of neighbouring threonine (hydroxylated) residues about the labelling site, and the presence of a number of repeated sequence motifs eg. -P-P-, and -V-V-V- such have been noted in other substrates (see section 1.8).

7.4. THE PREDICTION OF LOCAL SECONDARY STRUCTURE SURROUNDING GLN_{78}

Glutaminyl residue Gln_{78} is predicted by Holt and Sawyer, (1988) to lie in an unstructured region between a β-turn (extending from residue Pro_{70} through to Asn_{72}) and some β-sheet (residues Pro_{81} through to Val_{84}). Glutaminyl residue Gln_{71} is predicted to lie at the apex of this β-turn, and Gln_{88} is predicted to lie in the five residue connecting region between the aforementioned β-sheet and a region high in α-helix forming potential (ie. Glu_{91} through to Lys_{99}).

7.5. THE SPECIFICITY OF TRANSGLUTAMINASE TOWARDS GLUTAMINYL RESIDUES IN BOVINE β-CASEIN

The amino acid sequences surrounding the glutaminyl residues of bovine β-casein are presented in Table 1.3. and Table 1.4. (section 1.7.7.3.). Since the tertiary structure of β-caseins is not known it is not possible to say whether many of the glutaminyl residues are sufficiently exposed in native non-succinylated casein to be available to the active site of transglutaminase. However one glutaminyl residue is known to be highly

exposed, ie. Gln_{167} since this is the single site in native ß-casein known to be modified by factor XIIIa (Gorman an Folk, 1980). The sequence about this latter site is shown in Fig 6.5.

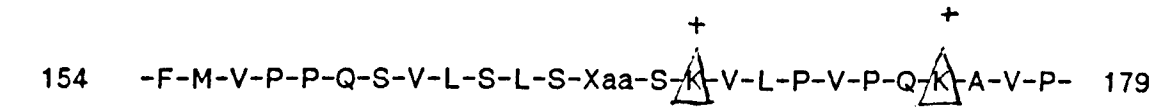


Fig 7.5. MAJOR SUBSTRATE SITE OF FACTOR XIIIa IN BOVINE ß-CASEIN [-Xaa- = N-ε(γ-glutamyl)dansylcadaverine] positively charged residues are in triangles and negatively charged residues are in circles

Database searches for sequences similar to that shown in Fig 7.5 have shown that the sequence surrounding Gln_{79} is one of the closest matches available (see section 9.3.3.2.) and for this reason Gln_{167} might be exp ected to have substrate properties for the liver transglutaminase. However in a number of other protein substrates of transglutaminase the presence of positively charged residues in regions adjacent to glutaminyl residues appears to reduce the substrate properties of these groups (see sections 1.7.8.4., 4.3.1., 5.3.3., and 6.3.3.. Therefore if Gln_{167} is reactive towards guinea pig liver transglutaminase it is unlikely on the basis of this labelling

study to be a major modification site due to the positive charge at position +2 imparted by Lys_{179}.

CHAPTER EIGHT: LESS WELL CHARACTERIZED SUBSTRATES OF GUINEA PIG LIVER TRANSGLUTAMINASE

8.1. GENERAL INTRODUCTION

A number of other protein and peptide substrates were determined to have substrate properties although the specific modification sites were not directly determined, but could be inferred from indirect evidence eg.

1) known tertiary structure,
2) maximum observed stoichiometry of labelling,
3) characteristics of the amino acid sequence of the substrate,
4) peptide mapping of labelled material.
5) partial sequence data inferred from peptide mixtures.

Some of these substrates are considered in sections 8.2. to 8.4..

8.2. CATALASE (BOVINE)

8.2.1.INTRODUCTION

Catalase has been previously noted as a possible substrate for transglutaminase (Lorand and Conrad, 1984). Catalase has been fully sequenced via a combination of enzymic, and chemical methodologies (Schroeder et al., 1982). The X-ray crystallographic structure is available to a resolution of 2.5 Å (Murthy

et al., 1981), where the known sequence has been unambiguously fitted to the electron density map. Catalase consists of four subunits in the crystal structure, but is known to dissociate in solution into dimers and monomers at high pH (Inada et al. 1961). On the basis of what is currently known concerning the specificity of transglutaminase the most likely sites of modification of bovine catalase have been predicted (see section 4.4.3.). These predictions have been further substantiated by proteolytically fingerprinting the modified protein using HPLC techniques.

8.2.2. METHODS

Bovine liver catalase, was purchased from Sigma as a lyophilized powder, and stock solutions were constructed using an extinction coefficient of 1.63 for a 1 mg/ml solution at 280 nm (Theorell, H. (1951). The assay used to asses the specific activity of catalase was that of Bergmeyer, (1955), see appendix II. The substrate properties of catalase for transglutaminase were tested in the absence of, and in the presence of 1.0 and 5.0 mM DTT. Transglutaminase catalysed modification was determined at 25°C, and at pH 7.6 in triethanolamine buffer (50 mM), and diluted into phosphate buffer prior to enzyme assay. Incorporation profiles of both putrescine and dansylcadaverine were constructed and the labelled sites were characterized by peptide mapping by HPLC. This was possible since the tryptic, and chymotryptic peptides have been extensively characterized (Schroeder et al. 1964).

8.2.3. RESULTS AND DISCUSSION

8.2.3.1. MODIFICATION OF CATALASE

The modification of catalase with [^{14}C]-putrescine and dansylcadaverine indicated that labelling was both Ca^{2+}, and transglutaminase dependent. Dansylcadaverine binds quite extensively and non-specifically to catalase up to a stoichiometry of ~ 1 - 2 mol amine/mol protein. Dansylcadaverine was released on proteolytic digestion since a fluorescent peak was observed at the position of the dansylcadaverine standard following HPLC. This made the calculation of the stoichiometry of labelling by dansylcadaverine incorporation prone to error. The maximum number of reactive glutaminyl sites within catalase was judged from the transglutaminase catalysed incorporation of putrescine, since non-covalently bound putrescine is easily removed from samples by the filter disc method, (see section 2.2.3.2.). Since between 2.0 and 3.0 mol putrescine/mol protein was incorporated into catalase after 20 hr then the maximum number of reactive sites is estimated to be three..

The incorporation of label appeared to be partially dependent on the presence of DTT. This was perhaps surprising, since no disulphide bridges which could be reduced and broken by DTT are apparent from the X-ray structure of catalase. It is possible that surface exposed cysteine residues within the catalase structure could form disulphide bonds with Cys_{275} at the active site of transglutaminase. This could cause inactivation of enzymic activity by blocking the regeneration of thioester intermediates in the catalytic mechanism (see section 1.5.). Cysteine residues which are exposed at the surface of catalase include Cys_{392}, and Cys_{459}.

8.2.3.2. SOLUTION STRUCTURE OF CATALASE

The crystal structure (Eventoff and Gurskaya, 1975) indicates that catalase is usually a tetramer but at high pH this dissociates into monomers (Inada et al., 1961). Hence the solution structure of catalase during transglutaminase labelling is expected to resemble the crystal structure shown in Fig 4.12.. The result of circular dichroism analysis of catalase in the far u.v. range (taken under conditions resembling those used during labelling) is shown in Fig 8.1.. The structure appeared to remain stable over 24 hr incubation at 20 °C. This was also indicated by the stability of the specific activity of catalase during labelling. A CONTIN (Provencher and Glöckner, 1981) analysis of the secondary structure of catalase in the labelling mixture gave the following results: [α-Helix 19 % +/- 2 %, β-sheet 36 +/- 2 %, and other 45 % +/- 2 %]. This is rather different from that calculated from the X-ray structure (Murthy et al., 1981) ie. [α-helix 26 %, β-sheet 12 %]. However these figures are closer to those derived from the X-ray structure than those previously obtained from optical rotatory dispersion (ORD) measurements. Yang and Samejima (1963), estimated bovine catalase to contain around 50 % helix using these methods.

8.2.3.3. PARTIAL CHARACTERIZATION OF SITES OF LABELLING

Catalase was heat treated and the pellet and supernatant were digested using bovine trypsin (general method of Schroeder et al., 1964). This was ineffective at producing peptide material which could be eluted under standard conditions from a Vydac C18 reverse phase column,

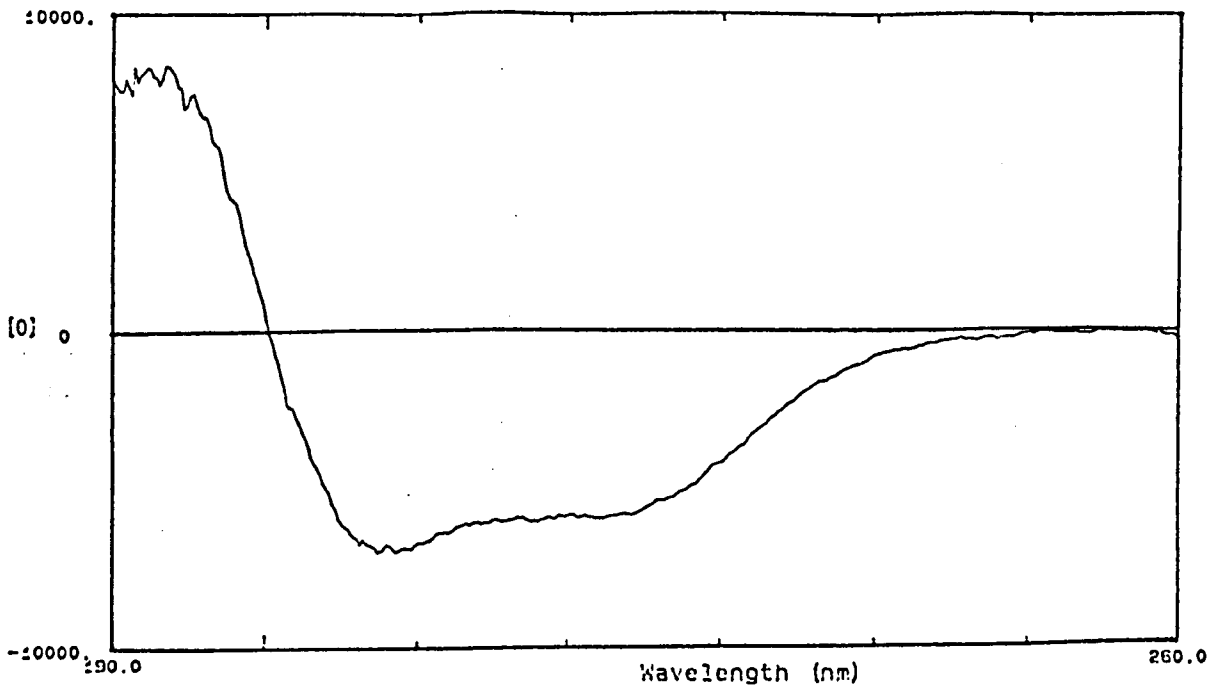


Fig 8.1. FAR U.V. CD SPECTRUM OF BOVINE LIVER CATALASE.

The protein concentration was 1.0 mg/ml and data was collected at 25 °C using the standard conditions described in section 2.2.4.3.. The secondary structure of catalase was calculated on the basis of this data using CONTIN analysis (Provencher and Glockner, 1981). The trace represent averaged structures of two background corrected scans.

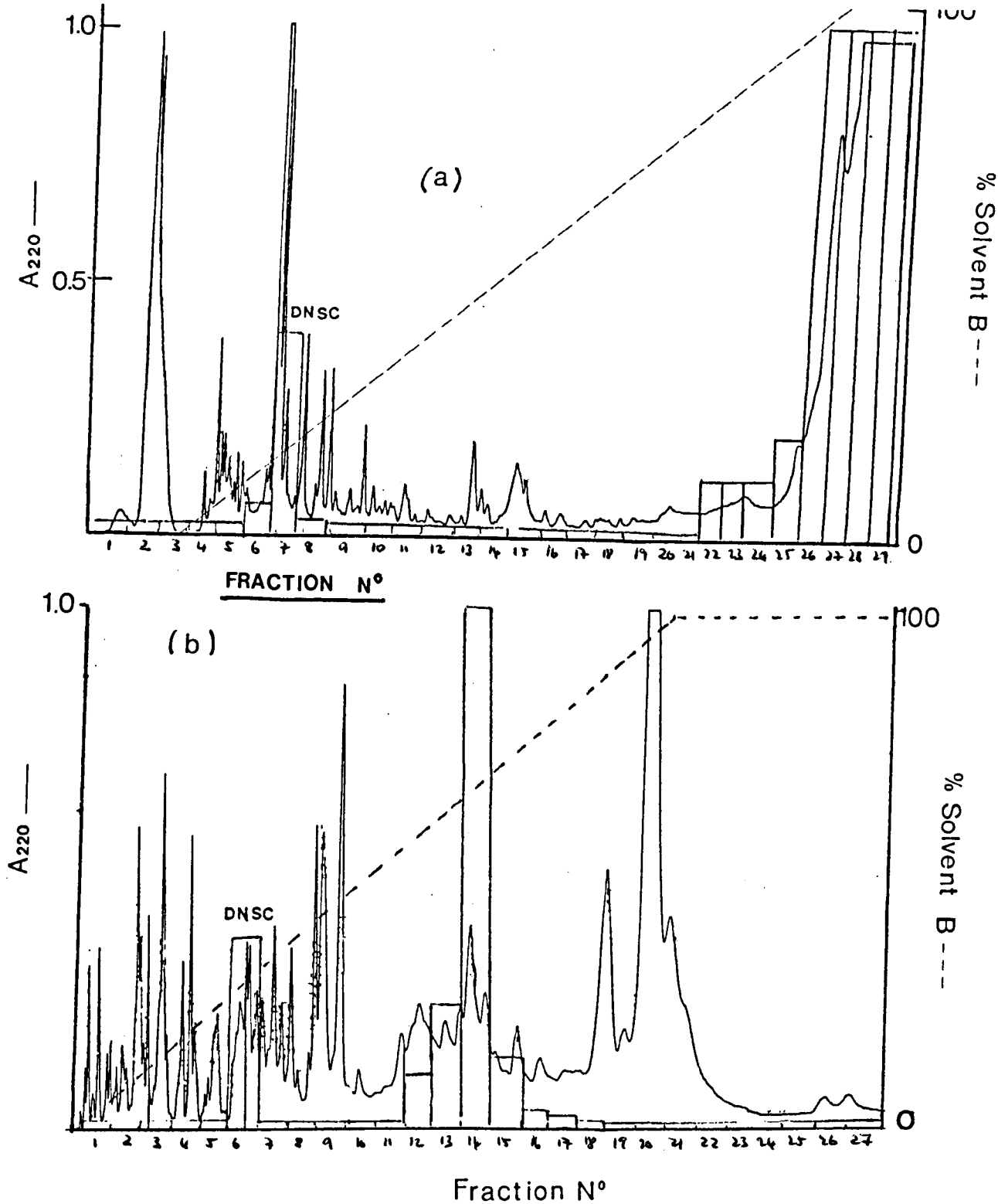


Fig 8.2. THE SEPARATION OF PROTEOLYTIC PEPTIDES OF CATALASE BY HPLC

(a) HPLC separation of the products of the Initial digestion of catalase (1/40 wt proteinase/wt catalase) using TPCK treated bovine pancreatic trypsin on heat treated catalase. The fluorescent material represented modified protein since unbound dansylcadaverine (standard) was shown to elute at t = 17 min on this gradient. The gradient in this case was 5.0 - 100 % B over 90 min, where B = 70 % isopropanol in water with 0.4 % (v/v) triethylamine phosphate pH 2.5, and A was water with 0.4 % (v/v) triethylamine phosphate pH 2.5.

(b) HPLC of the tryptic digestion of carboxymethylated catalase (method of Allen, 1981) on the same gradient as (a) The addition of chymotrypsin directly into the mixture which was digested with trypsin, produced fluorescent peptides in similar yields to that obtained in Fig (c) though of course the trace was more crowded by peptide peaks. The digestion of guanidine hydrochloride treated, non-carboxymethylated catalase gave a trace similar to that in (a) indicating that the slow dialysis of denaturant facilitated refolding of the protein.

The rectangles represent the relative fluoresecnce associated with 3ml fractions of column eluate.

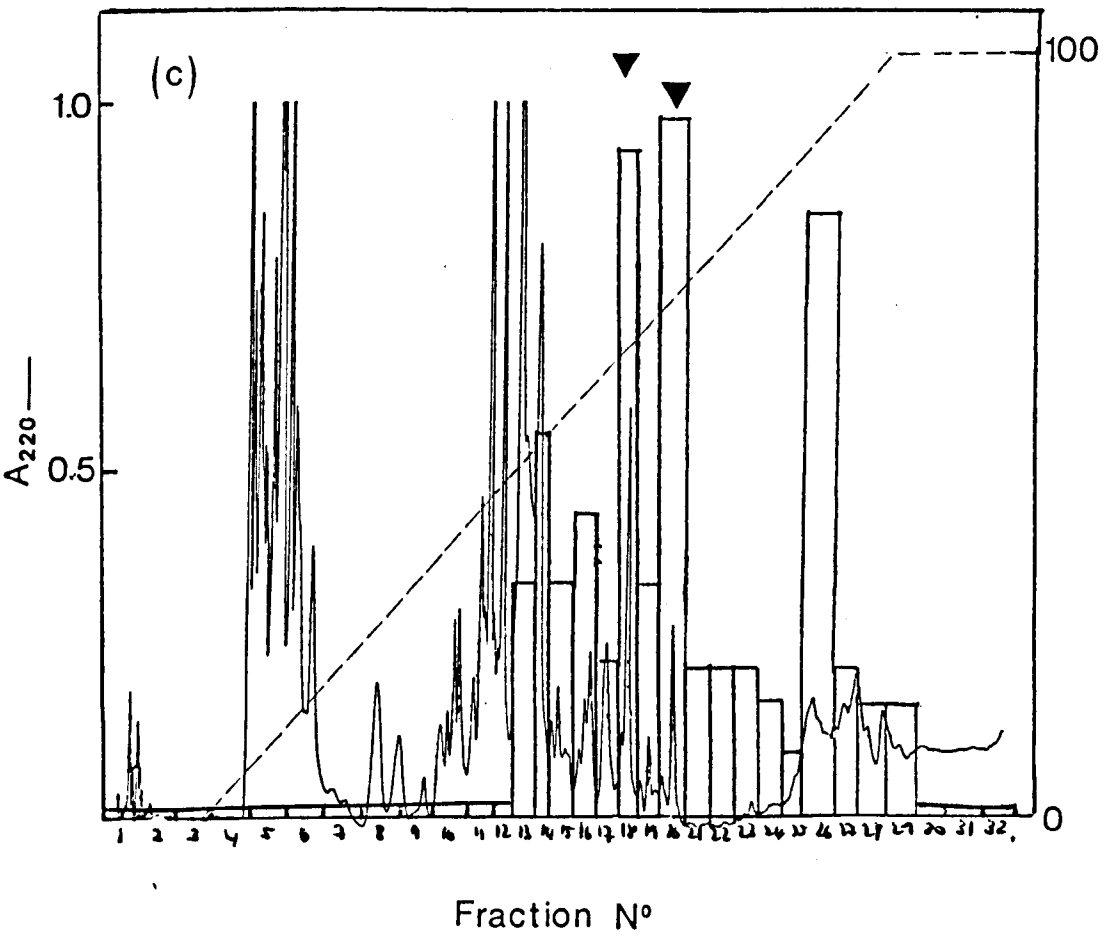


Fig 8.2. THE SEPARATION OF PROTEOLYTIC PEPTIDES OF CATALASE BY HPLC

(c) HPLC trace of the products of the chymotryptic digestion of the major dansylated peptide from trace (a), using the same gradient as in (a) but over 60 min. The material was firstly purified by gel filtration on a sephacryl G75 column and then reduced in volume from 5ml to 0.5 ml by rotary evaporation. 0.5 ml of 2 % (w/v) ammonium bicarbonate solution was added to give a pH of 8.0 whereupon 1/40 wt proteinase/wt catalase peptide of TCLK bovine pancreatic chymotrypsin was added. The incubation was terminated after 4.0 hr by injection onto the HPLC column. The trace stayed unchanged when digestion was continued up to 24 hr. The gradient was the same as that used in(a). After rechromatography the two most fluorescent peaks▼were collected and sent for sequence analysis to Celltech Ltd, Slough.

even with high concentrations (> 40 % v/v) of isopropanol. Repeated additions of trypsin and up to 24 hr incubation did not produce sufficient material for analysis. The peptide(s) which contain the label were either very large (due to incomplete digestion) and/or hydrophobic. The same 24 hr digestion of [^{14}C]-putrescine labelled material which had been carboxymethylated in its denatured form (ie. in the presence of 6 M guanidine hydrochloride) was more successfully digested, and a single major fluorescently labelled peptide was observed to elute from the column at approximately 35 % v/v isopropanol (see Fig 8.2 b.).

In order to produce modified peptides suitable for sequencing it was necessary to sub-digest the tryptic peptide which contained most of the label. This was could be achieved by the purification of the peptide by size exclusion chromatography using a Sepadex G-75 column (dimensions 30 cm x 3.5 cm) equilibriated with 1 % w/v ammonium bicarbonate pH 8.0, followed by digestion with bovine chymotrypsin at a ratio of 1/40 wt/wt catalase. Alternatively the chymotrypsin could simply be added into the tryptic digestion mixture in the same ratio and the products separated by HPLC (see Fig 8.2 (c)). Unfortunately direct identification of the modification sites within catalase was not possible since purified peptides appeared to stick to the plastic eppendorf tubes in which they were stored. However on the basis of features common to other substrates of transglutaminase it is possible to predict which glutaminyl residues are most likely to have substrate properties (see section 4.4.3.). The external N-terminal tail appears an unlikely site for modification due to the high concentration of positive charge which is associated with the sequences containing glutaminyl residues in this region.

8.3. PEPSIN (PORCINE)

8.3.1. INTRODUCTION

Porcine pepsin was prepared as a 10 mg/ml stock solution in Tris-HCl (100 mM) buffer pH 8.0, and initially incubated at 4°C for 2 hr. The enzyme is catalytically inactive at this pH and no autolysis of the Mr 34 kDa band corresponding to pepsin was observed by SDS-PAGE. The stock solution was then diluted into standard incubation buffer used for protein labelling [minus β-casein] (see section 2.2.3.2.) and tested for substrate properties using both putrescine and dansylcadaverine as acyl acceptors, both in the presence and absence of DTT over the range from pH 6.0 to pH 9.0.

Pepsin which had been dansylated for 10 hr at pH 7.5 was enzymically digested with *Staphylococcus aureus* V8 proteinase and the peptide products separated using an electrophoretic technique on an 8 to 25 % gradient polyacrylamide gel which contained both SDS and urea. This was electroblotted onto an Immobilon sheet (see section 2.2.4.4.) and the labelled peptides visualized under u.v. light. The most highly fluorescent band was excised and prepared for sequence determination.

Circular dichroism studies were conducted on pepsin at pH 7.6 both in the presence and absence of 1.0 mM DTT. Spectra were obtained using 50 mM Tris-HCl buffer (see section 2.2.4.3.). The effect of Ca^{2+} on the secondary structure of pepsin was tested both in the presence and absence of DTT.

8.3.2. RESULTS AND DISCUSSION

8.3.2.1. C.D. STUDIES

The far u.v. spectrum of pepsin at pH 7.6 is shown in Fig 8.3 .

A CONTIN (Provencher and Glöckner, 1981) analysis indicated that the protein comprised the following secondary structure at pH 8.0 ,[α-helix: 11 % +/- 0.5 %, ß-sheet: 44 % +/- 1 %, and other: 45 % +/- 1 %]. Addition of DTT appeared to have little effect on the spectra after 1 hr incubation. After 14 hr incubation at 20 °C CONTIN analysis indicated a small increase in helical content [α-helix: 14 % +/- 0.4 %, ß-sheet: 44 % +/- 1 %]. However subsequent addition of 5 mM Ca^{2+} to the overnight mixture resulted in a rapid change in the spectra. CONTIN analysis indicated that the change derived largely from conversion of α-helix to ß-sheet structure [α-helix: 2 % +/- 1.4 %, ß-sheet: 58 % +/-3 %, and other: 40 % +/- 2 %]. Such a change was not observed when calcium ions were present, but DTT was not included in the mixture. The results presented here were comparable with the secondary structural content of pepsin determined by X-ray crystallography (Andreeva et al., 1978). This is estimated to comprise : 10 % α-helix, ß-sheet 33 %, and other 57 %.

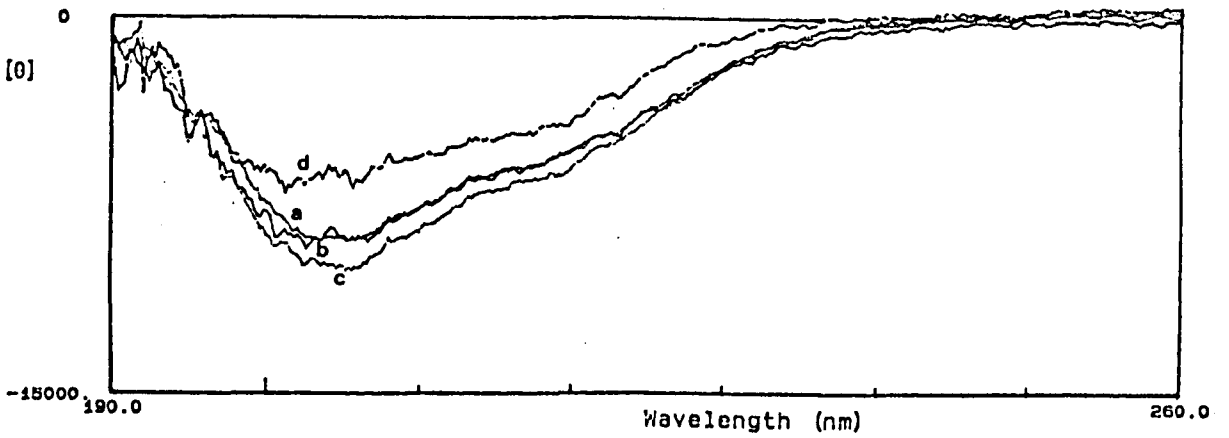


Fig 8.3. FAR U.V. CD SPECTRUM OF PORCINE PEPSIN.

A concentration of 1.0 mg/ml of pepsin was used. Samples were incubated at 25°C under standard conditions for 2 h prior to their analysis. A CONTIN analysis was done on the data in order to predict the amount of secondary structure within the protein under the stated conditions. Each spectrum is an average of two blank corrected scans.

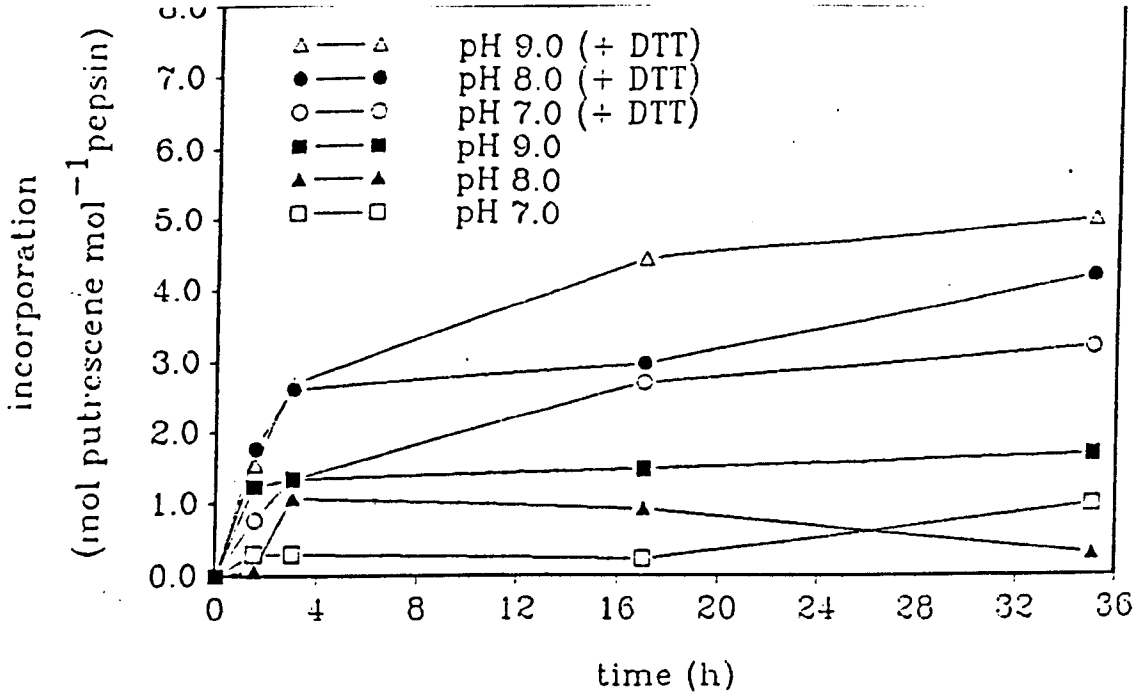
(a) In tris-HCl buffer pH 8.0.
(b) In the presence of 1 mM DTT
(c) In the presence of 1 mM DTT overnight
(d) after adding Ca^{2+} to (c)

Both the near and far u.v. spectrum of porcine pepsin recorded at pH 8.0 was very similar to that at pH 2.0 and hence the protein is thought to resemble its crystal structure under the conditions used for labelling by transglutaminase.

8.3.2.2. LABELLING STUDIES

The time dependent incorporation of putrescine into pepsin at a variety of pH values is illustrated by Fig 8.4. The presence of 1.0 mM DTT has a large effect on the maximum observed stoichiometry of labelling over 22 hr. ie. in the presence of DTT up to 5.0 - 6.0 mol putrescine was incorporated per mol pepsin monomer, compared to only 2.0 mol putrescine/mol porcine pepsin in the absence of reducing agent. Similar results were obtained using dansylcadaverine as the amine substrate. The maximum observed stoichiometry following two additions of transglutaminase to the mixture, and incubating for 36 hr in the presence of 5 mM DTT was 5.0 mol DNSC/mol pepsin.

The difference in the maximum stoichiometry of labelling in the presence of DTT may be correlated with the loss of secondary structure associated with the reduction of disulphide bridges. The effect of the DTT appeared to be potentiated by the action of Ca^{2+} binding to the substrate. The negatively charged amino acid side chains in pepsin, may be electrostatically attracted to positively charged calcium ions and this interaction may accelerate the denaturation of pepsin under labelling conditions. The substrate properties of pepsin are discussed in relation to pepsin structure in Appendix xi.



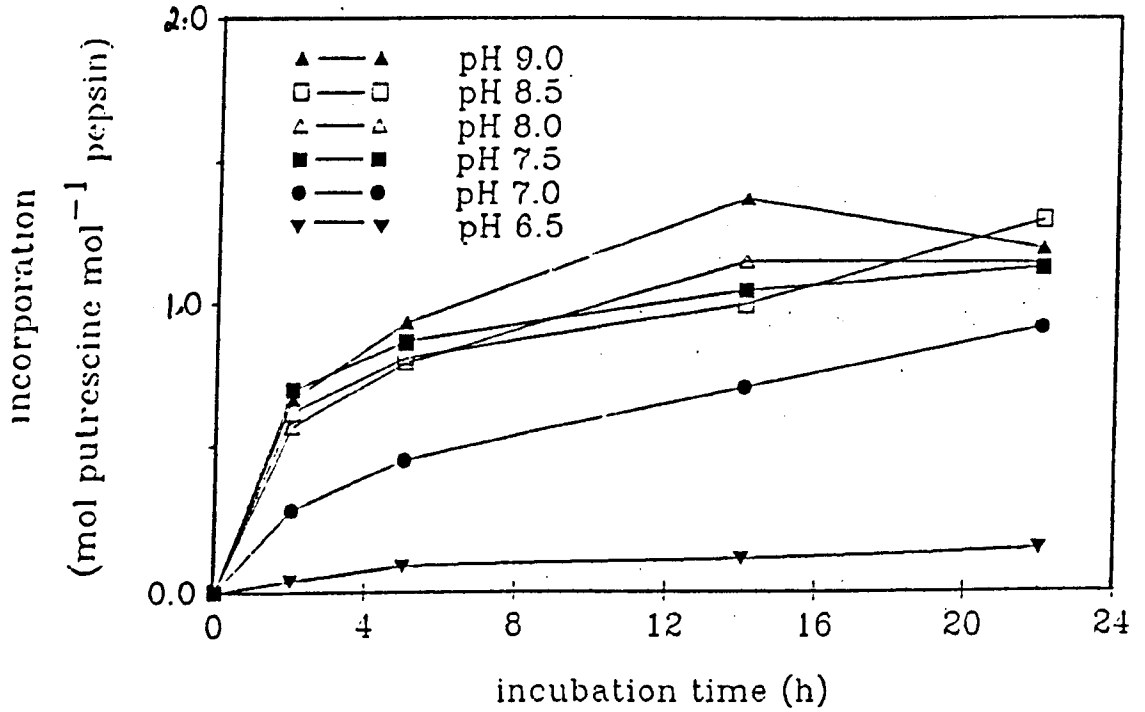


Fig 8.4. THE INCORPORATION OF PUTRESCINE INTO PEPSIN BY TRANSGLUTAMINASE

The modification of pepsin was under the standard conditions described in section 2.2.3.2.. The protein concentration was 3 mg/ml. Samples were preincubated at pH 8.0 in 25 mM tris HCl buffer (under which conditions the majority of tertiary structure is retained, see Fig 8.5.) and then the pH was adjusted to the required pH using 100 mM buffer of the desired pH. Final adjustments of pH were made using dropwise addition of dil HCl or NaOH (aq). with incubated at 25 °C. No proteinase activity was detected in any of the samples tested and little or no cross-linking of pepsin monomers was observed using SDS PAGE during the reaction with transglutaminase. No incorporation of putrescine was observed in the absence of transglutaminase.

8.3.2.3. DETERMINATION OF A PUTATIVE MODIFICATION SITE WITHIN PEPSIN

Peptide sequencing of an electroblotted V_8 fragment of dansylated pepsin prepared in the absence of DTT provided evidence for the sequence:

V-F-I-R-X-Y-Y-T

No peak corresponding to Gln_{308} was observed (X). This is typical of the behaviour of dansylated glutamyl derivatives during sequencing and has been noted elsewhere (see section 5.3.1. and 6.3.1.). This sequence contains features observed in other transglutaminase substrates ie.

1) N-terminal positive charge neighbouring a reactive Gln,
2) high hydrophobicity,
3) high content of hydroxylated amino acids,
4) is located within the C- terminal section of a protein substrate, as seen in the case of β-lactoglobulin, (see section 5.3.2.).

The location of glutaminyl residues within the tertiary crystal structure of porcine pepsin is shown in Fig 8.5. The solution structure of pepsin under the conditions used for labelling of pepsin (pH 8.0) is expected to differ somewhat from this structure particularly following reduction of disulphide bonds. The similarity of the tertiary structure is however considerably greater in the absence of DTT than in its presence (see section 8.3.2.1.).

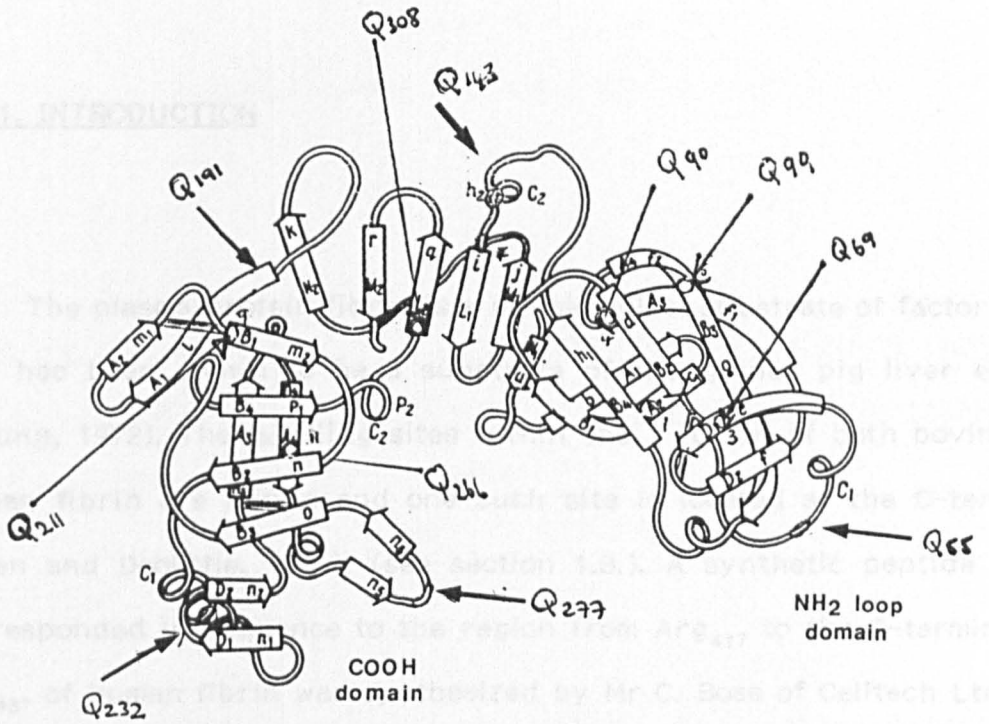


Fig 8.5 THE X-RAY STRUCTURE OF PORCINE PEPSIN

Schematic drawing of the trace of the polypeptide in porcine pepsin. Regions of ß-sheet are arrow-like bars and α-helix are shown as multiple turns within the remaining strands. The different strands of ß-sheet are alphabetically labelled according to the notation of Andreeva and Gustchina (1979). The approximate location of some of the glutaminyl residues are indicated by arrows. From c.d. studies of the protein it appears that even at pH 8.0 and in the presence of the components of the transglutaminase incubation mixture (ie DTT) pepsin still resembles this structure and is not denatured despite having a very low pH maximum for its enzyme activity.

8.4. A SYNTHETIC FIBRIN PEPTIDE AS TRANSGLUTAMINASE SUBSTRATES

8.4.1. INTRODUCTION

The plasma protein fibrin is a physiological substrate of factor XIIIa, and has been shown to be a substrate of the guinea pig liver enzyme (Chung, 1972). The labelling sites within the γ-chain of both bovine and human fibrin are known and one such site is located at the C-terminus (Chen and Doolittle, 1971), (see section 1.8.). A synthetic peptide which corresponded in sequence to the region from Arg_{417} to the C-terminus at Tyr_{437} of human fibrin was synthesized by Mr C. Bose of Celltech Ltd. The sequence is shown in Fig 8.6.

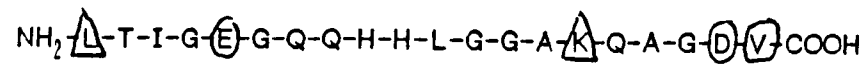


Fig 8.6. THE SEQUENCE OF A SYNTHETIC PEPTIDE BASED ON FIBRIN positively charged residues are in triangles an negatively charged residues are in circles

The secondary structure of this region of fibrin is not known, but has been modelled as a region of α-helix by Doolittle, (1973) [see appendix I]. Recently an nmr study of a peptide of similar sequence to that shown in Fig 8.5 indicated that below pH 5.0 α-helical structure may be associated with this region of fibrin, (Mayo and Burke, 1990). However above this pH resonance intensities were greatly attenuated due to increased rates of exchange.

The substrate properties of this peptide were investigated in order to test the feasibility of constructing a cross-linked dimer using transglutaminase. Cross-linked material was then tested for its ability to cause antibody production in laboratory animals. An antibody raied against such an antigen could theoretically be of use pharmacologically if it were conjugated to proteinases as a thrombolytic agent.

8.4.2. METHODS

The peptide was easily solubilized in standard transglutaminase protein labelling buffers eg. tris-HCl (50 mM) pH 7.6 up to concentrations of 1.0 - 3.0 mg/ml. Transglutaminase cross-linking mixtures were of the same basic composition as standard labelling mixtures (see section 2.2.3.2.), except that the primary amine which was not included and the peptide replaced β-casein in the mixture. The time course of the modification of this peptide by transglutaminase was monitored by HPLC, and TLC. Cross-linked material was analyzed by F.A.B. mass spectroscopy.

The solution structure of the material was analyzed by circular dichroism, both in normal aqueous solution at pH 7.6 and at pH 3.0 in both the presence and absence of trifluoroethanol. This was compared to a computer generated secondary structure prediction of the peptide, which was conducted by Dr L. Sawyer, Department of Biochemistry, University of Edinburgh. The peptide was modelled using computer graphical techniques (see section 2.2.8.1.) into a variety of conformations in order to test hypotheses concerning the secondary structure of the region of polypeptide chain in the vicinity of the labelling sites.

8.4.3. RESULTS AND DISCUSSION

8.4.3.1. THE SECONDARY STRUCTURE OF FIBRIN PEPTIDE

a) CIRCULAR DICHROISM STUDIES

A CONTIN (Provencher and Glöckner, 1981) analysis of the far u.v. spectrum of the fibrin peptide (not shown) indicated the presence of some secondary structure at pH 7.6 and this was apparently retained at pH 3. The peptide was found to contain 57 % +/- 1 % β-sheet. No α-helix was present. The addition of Ca^{2+} and TFE (5 % v/v) did not appear to affect the spectrum. It is possible that the peptide consists of a series of β-turns and that these structures contribute towards the β-sheet content.

b) SECONDARY STRUCTURAL PREDICTION OF PEPTIDE STRUCTURE

The peptide was predicted to contain little secondary structure according to the programme PREDICT [see section 2.2.8.2.]. Individual algorithms suggested that the peptide may contain some β-turn potential. In no cases did any of the programmes tested predict an α-helical structure within this sequence.

c) MODELLING OF THE CONFORMATION OF FIBRIN PEPTIDE

This sequence has previously been modelled by Doolittle, (1973) as an α-helix, and as a β-structure, see Fig 8.6. The CHARMm programme was used to model the sequence into a number of conformations by an energy minimization process. The programme suggested that both a negative or positive α-helical conformation was energetically more favoured than the extended conformation.

8.4.3.2. MODIFICATION OF FIBRIN PEPTIDE

A time course of incorporation of label into the fibrin peptide was obtained by injecting known amounts of peptide from standard ^{14}C-putrescine containing labelling mixtures onto the vydac HPLC column and monitoring the incorporation of amine by scintillation counting (the column conditions immediately quenched the activity of transglutaminase). The results indicated that the peptide was a highly efficient substrate for transglutaminase, as has been previously been reported for the intact molecule (Chung, 1972), the derivatized γ-chain (Gorman and Folk, 1980), and synthetic peptides based on the amino acid sequence surrounding the Factor XIIIa cross-linking site within fibrin (Gorman and Folk, 1980). The maximum stoichiometry of labelling observed was 2.2 mol putrescine/mol peptide and was achieved after 2.0 hr of incubation. No reaction was detected in the absence of Ca^{2+} or transglutaminase. The omission of DTT had no observable effect on the time course of reaction. The HPLC traces associated with these modifications are shown in Fig 8.8 i.

The omission of primary amine in the incubation mixtures resulted in the rapid formation of a number of novel peaks when the components of the mixture were analyzed by reverse phase HPLC. The time course for the production of this product is shown in Fig 8.8 ii. Since the peptide had already been shown to contain reactive glutaminyl residues, it was assumed that the new peaks represented cross-linked product (the lysine residue at position Lys_{14} acting as amine donor). Approximately 2 mg of modified sample was sent to Celltech Ltd. for further characterization by FAB mass spectrometry and for use as an antigen in trials to produce antibody against cross-linked material.

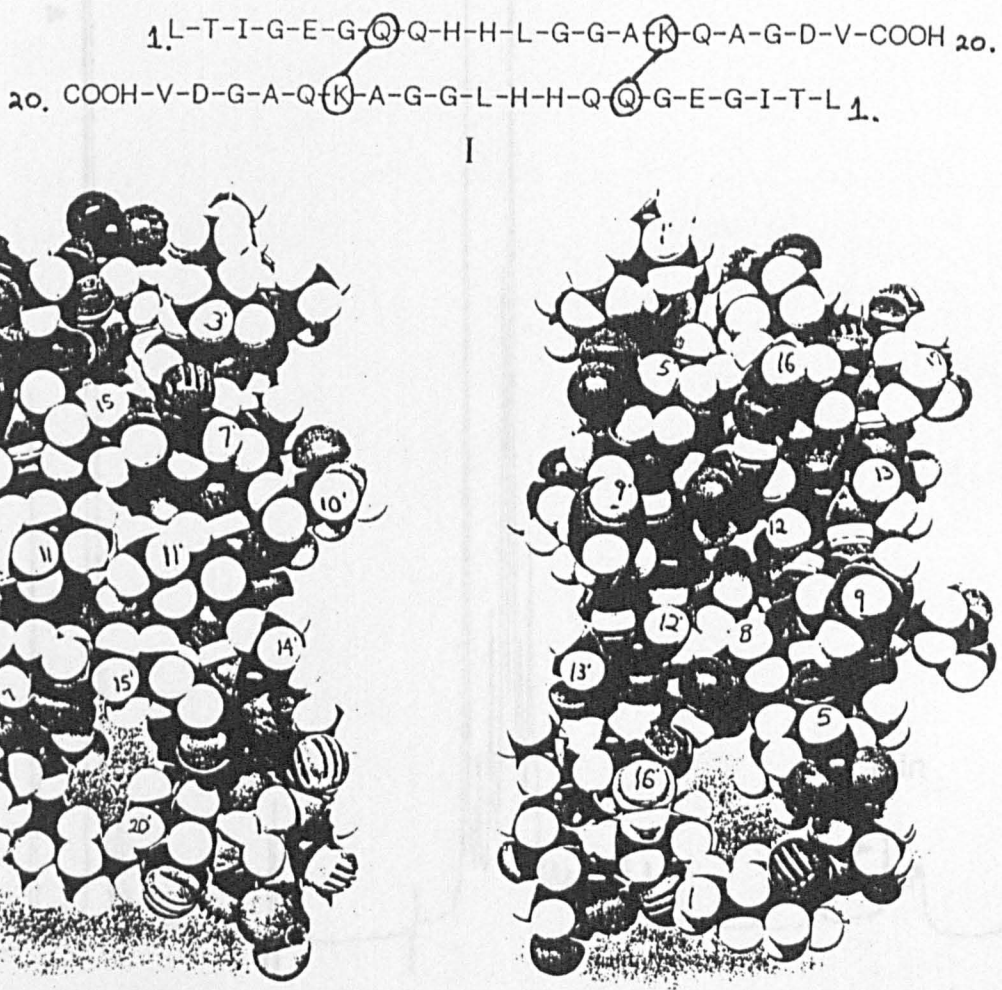


Fig 8.7.

AN α-HELICAL MODEL OF THE C-TERMINUS OF HUMAN FIBRIN

I) The amino acid sequence of the carboxy termini of the fibrin γ-chains showing the locations of reciprocal cross-links between antiparallel neighbouring chains.(o-o).

II) A space filling models of the γ- chain carboxy-terminal eicosopeptides in an α-helical conformation, orientated in antiparallel register. (a) Shows the cross-links formed between Lys_{15} and Gln_7. (b) Shows the model from the opposite side, showing the possible interaction between Gln_8 and Cln_{16}. The diagram was taken from Doolittle (1973).

These peptide models conform exactly in sequence to that of the peptide used for the experimental work done as part of this thesis (see section 8.4). It is interesting to speculate whether the conformation of the peptide shown above may be that which transglutaminase binds during its modification. Certainly other protein substrates are predicted to have helical structure eg. phospholipase A2, lipocortin I.

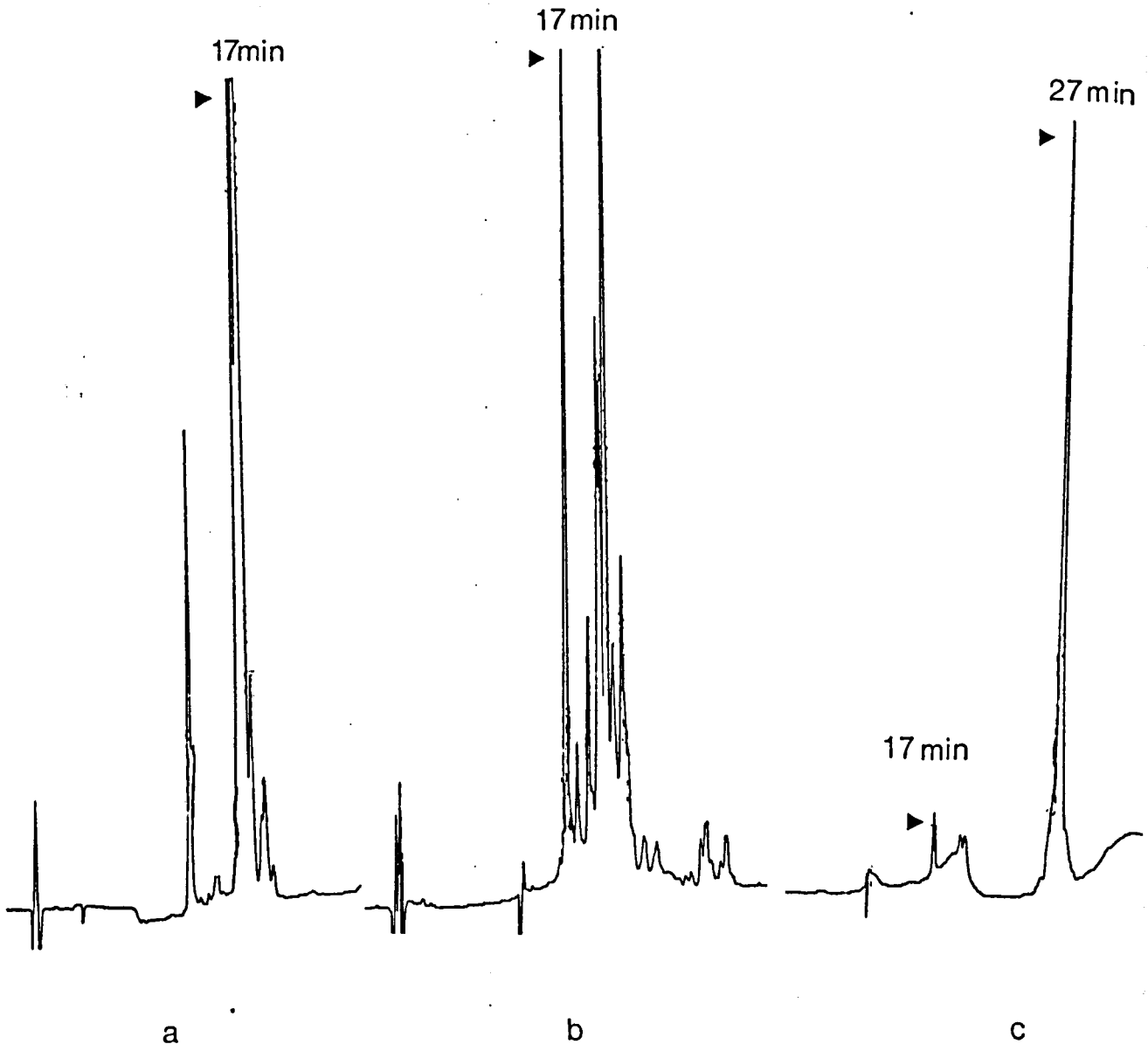


Fig 8.8. HPLC MAPPING OF THE PRODUCTS OF TRANSGLUTAMINASE MODIFICATION OF THE FIBRIN PEPTIDE

Samples were incubated according to the text and 0.1 mg of peptide was applied to a Vydac column after different times. The samples were eluted using a 10 - 70 % gradient of isopropanol in water (0.4 % v/v triethylamine phosphate pH 2.5) over 35 minutes.

i) In the presence of amine the modified peptide was converted to a number of radioactive forms (approximately 2.5 mol amine/mol peptide), which did not change position compared to the control (minus transglutaminase) over 24h. When dansylcadaverine is used instead of putrescine the peak shifted from t = 17 min to t = 27 min and approximately 1.5 mol amine/mol peptide was calculated to have been bound. Trace (a) is the control, trace (b) the putrescine modified peptide, and trace (c) the dansylated peptide.

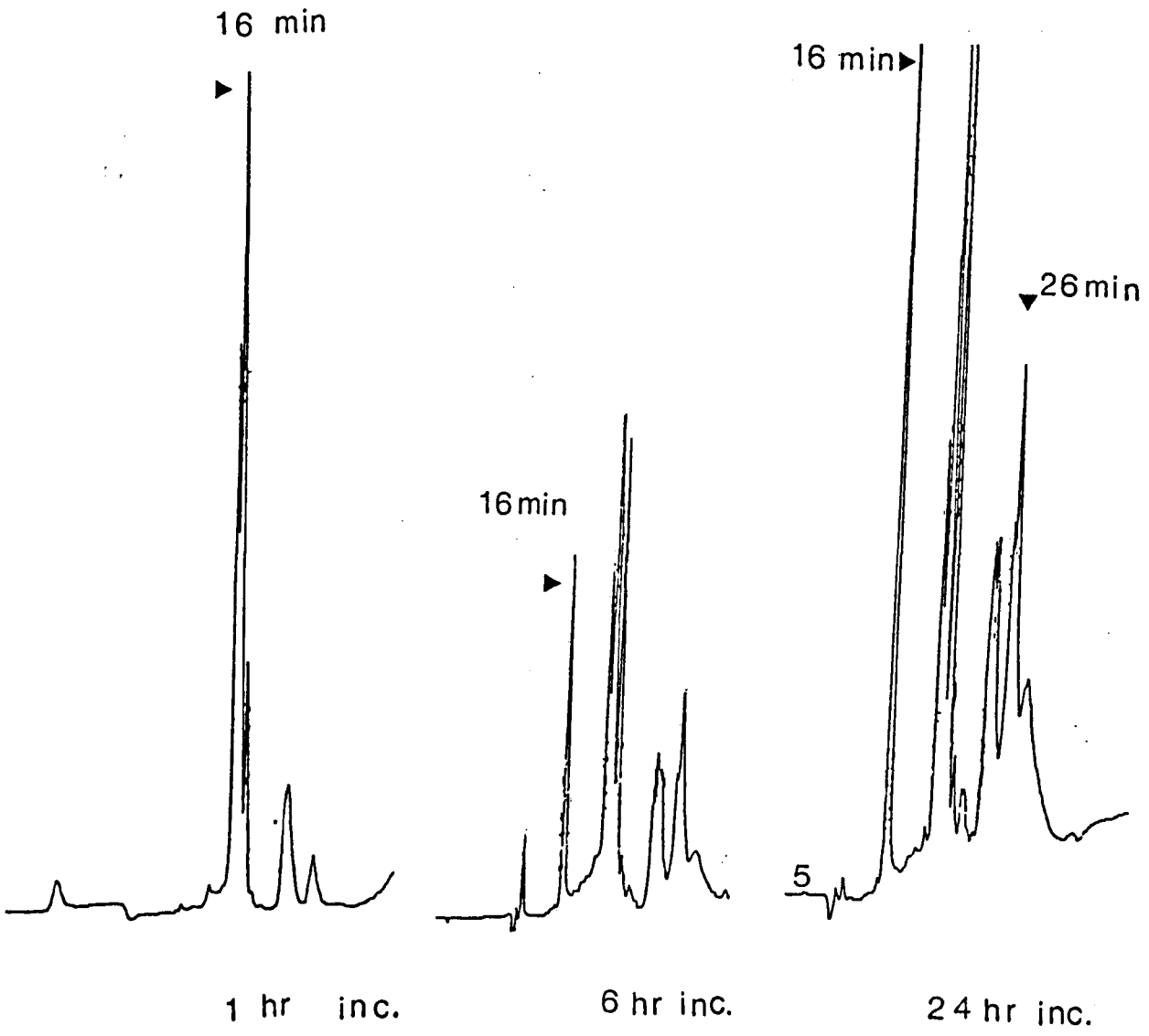


Fig 8.8. HPLC MAPPING OF THE PRODUCTS OF TRANSGLUTAMINASE MODIFICATION OF THE FIBRIN PEPTIDE

Cross-linking mixtures were constructed in the same way as in (Fig 8.8 i) with the exception that amine was excluded from the mixture. The HPLC separations were also done using the same gradient.

ii) In the absence of amine substrate a number of new peaks which elute at higher solvent concentrations are generated during the 24 hour time course. This is illustrated by the traces shown above.

8.4.3.3. PREDICTION OF REACTIVE GLUTAMINYL RESIDUES

Previous studies on related peptides and intact fibrin have shown that although both glutamines in the -Q-Q- motif are modified by transglutaminases, modification of each of the doublet tends to be less than 100 % probably due to steric restriction about the unmodified glutamine following the modification of the first in the series (Gorman and Folk, 1980). Thus although the sites of modification were not directly determined by sequencing of the labelled material prepared here it is likely that all three glutaminyl residues in the sequence have substrate properties for guinea pig liver transglutaminase.

8.4.3.4. PRODUCTION OF A CROSS-LINKED PEPTIDE ANTIGEN

The fibrin peptide was an efficient substrate of transglutaminase in the absence of free primary amine substrate (see Fig 8.8 ii),since a number of new peaks were observed to form, in a time dependent manner following HPLC separation of reaction mixtures. These novel peaks were not observed in the absence of transglutaminase. However in the absence of free primary amine it appears that contrary to catalysing cross-linking reactions, that transglutaminase may act by deaminating the glutaminyl residues of the peptide to glutamic acid. This reaction appears to predominate despite the high concentration of lysine residues in reaction mixtures. If this were the case the peptide would be expected to have a

high negative charge and may have interacted with the Ca^{2+} ions in solution to generate further peaks on the HPLC. The FAB mass spectrometric analysis suggested that only approximately 10 % of the sample was present as material of molecular weight corresponding to the dimeric form of the peptide (see Fig 8.9).

The suggestion has been made that the reciprocally linked segments may be analogous to a large cyclic (n=16) peptide (Doolittle, 1973). Such structures frequently employ reciprocal backbone hydrogen bonding for stabilization (Mosher and Blout, 1971). In view of this it may be significant that the peptide appears to have ß-sheet forming potential in solution [see section 8.4.3.1.(a)]. Whichever conformation is favoured by this sequence of amino acids in fibrin, it is necessary that the two strands align in such a way that a symmetrical dimer is the product of reaction.

8.4.3.5. PRODUCTION OF AN ANTIBODY AGAINST PEPTIDE CROSS-LINKED BY TRANSGLUTAMINASE

In unpublished experiments undertaken at Celltech Ltd., Slough [reported verbally by Dr B Smith], the ability of the cross-linked material to elicit anti-fibrin peptide antibody production in laboratory animals was investigated. The material did not appear to be very effective as an antigen and little antibody appeared to be produced in response to its injection into laboratory animals. It is possible that any cross-linked products were quickly sequestered from the bloodstream of the lab animals, or that the cross linked products were not very antigenic.

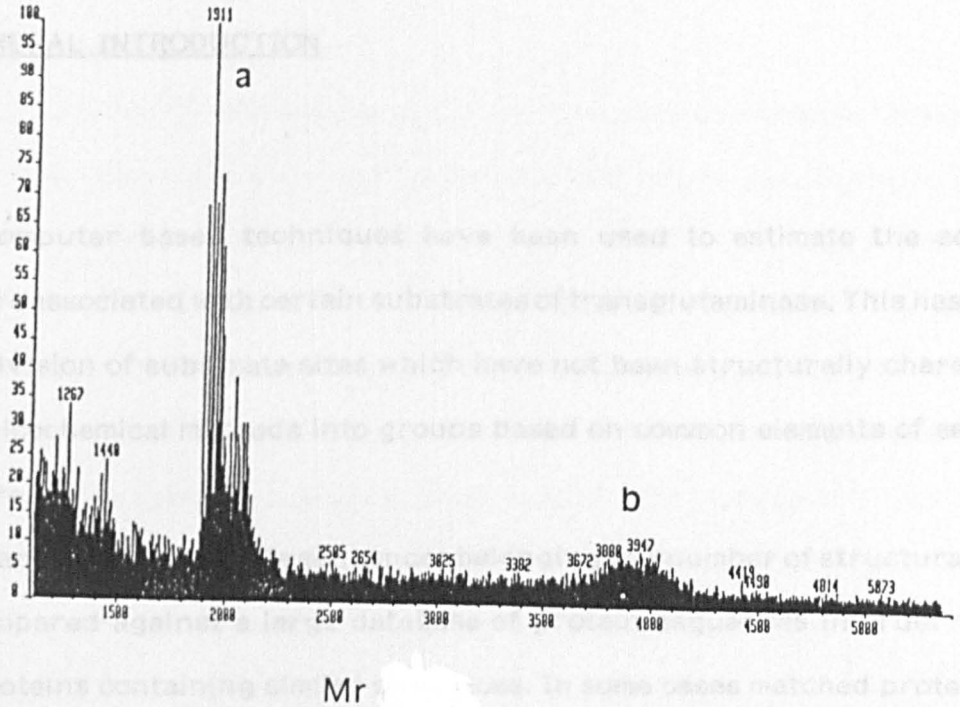


Fig 8.9. FAB MASS SPECTRA OF CROSS LINKING MIXTURE OF FIBRIN PEPTIDE.

The spectrum of material which was incubated in cross-linking mixtures (standard mixture minus exogenously added amine). The sample tested corresponds to that shown in Fig 8.8 following incubation for 24 hours in the presence of transglutaminase. The peak (a) at around Mr 1911 probably corresponds to that of the peptide and the peak (b) at Mr 3947 probably corresponds to the formation of some dimeric cross-linked material. The presence of other components of the incubation mixture eg. DTT, Ca^{2+} ions and buffer will have contributed to the noise in the trace at lower Mr values. From the change in the HPLC traces during incubation with transglutaminase it appears likely that major structural changes other than cross-linking occurred during the reaction. One possible reason for this phenomenon may have been the deamination of peptide by water in the absence of high concentrations of competing amine (see Fig 8.8),

CHAPTER NINE: COMPUTER BASED TECHNIQUES FOR THE STUDY OF TRANSGLUTAMINASE AND ITS SUBSTRATES

9.1. GENERAL INTRODUCTION

Computer based techniques have been used to estimate the secondary structure associated with certain substrates of transglutaminase. This has assisted in the division of substrate sites which have not been structurally characterised by physicochemical methods into groups based on common elements of secondary structure.

Many of the substrate sequences belonging to a number of structural classes were compared against a large database of protein sequences in order to locate other proteins containing similar sequences. In some cases matched proteins have previously been tested for their substrate properties. This approach provided a means for searching for structural consensus in the substrates of transglutaminase.

9.2. SECONDARY STRUCTURAL PREDICTIONS OF TRANSGLUTAMINASE SUBSTRATES

9.2.1. INTRODUCTION

Many algorithms have been developed to predict the local conformation of the chain of polypeptides and proteins from local sequence. The rationale for this work stems from experimental work which suggests that synthetic homopolymers

preferentially adopt certain conformations. In addition there is the notion that α-helices and β-strands and turns may act as nucleation sites for protein folding *in vivo*. The justification of this work is that better than random predictions can be generally obtained.

Methods may be divided into two types. The empirical schemes use parameters obtained from the analysis of known sequences and structures. The other approach is based on stereochemistry. Three commonly used methods are the empirical methods of Chou and Fasman (1974) and of Robson and coworkers (Garnier et al., 1978), and the stereochemical method of Lim (1974). These algorithms may be applied to a sequence and the results of each prediction statistically weighted, the results can then be combined to give a final result. This approach (built into the programme PREDICT) has been shown to have greater average predictive power than one single technique over a range of test sequences of known structure (see section 9.2.2.).

9.2.2. METHODS

Secondary structural predictions of protein and peptide structures were done by Dr L. Sawyer, Department of Biochemistry, University of Edinburgh. The predictions were in some cases done on relatively short sequences and for this reason a bank of eight different predictive methods were used and averaged in order to give the most likely conformation. The programme used was PREDICT (Eliopoulos et al., 1982) as modified by E. Eliopoulos (see Sawyer et al., 1986). For general background to this technique see section 2.2.8.2.

By using these techniques to estimate the secondary structure associated with substrate sequences of known three dimensional structure eg. actin,

phospholipase A_2, glucagon etc., it was possible to test the accuracy of the secondary structural predictions in these cases.

9.2.3. RESULTS AND DISCUSSION

Since only a limited number of protein substrates of transglutaminase have been characterized fully, the accuracy of the PREDICT programme must be assessed on the basis of a limited data set. However for those protein substrates tested the results of the programme concurred well with the secondary structure determined for these sites using more direct means eg. X-ray crystallographically determined structures. All three major classes of secondary structure ie. β-sheet, α-helix, and random coil structure were correctly associated with the appropriate regions of sequence within the substrate sites. The results are summarized in Table 9.1.. The amino acid sequences associated with the substrate sites of transglutaminase within protein and polypeptide substrates of unknown secondary structure were also analyzed by the PREDICT programme. The results are summarized in Table 9.2..

Secondary structural predictions of transglutaminase substrate sites are generally in agreement with the results of models constructed on the basis of physicochemical data. Substrate sites appear to include; helices (phospholipase A2 and lipocortin 1), loop structures (actin and PGK $His_{388} \longrightarrow Gln_{388}$ mutant), and a number of less folded structures higher in β-sheet and β-turn potential (β-casein, gliadins etc.).

In non-physiological substrates, higher secondary structural potential is generally not predicted. This was not always the case for physiological substrates

TABLE 9.1

The use of the computer program "PREDICT" to predict the known secondary structures associated with protein and polypeptide substrates of transglutaminase.

PROTEIN	SEQUENCE	PREDICTED STRUCTURE	KNOWN STRUCTURE
ACT	CRPRHQGVMVGMG	BBBTTTTTBBBB-	14 RESIDUE LOOP
GLC1	---HSQGTFTSDY	---TTTTT--TTT	FLEXIBLE REGION
GLC2	DSRRAQDFVQWLM	HHHHHHHHHHHHH-	HELICAL REGION
PLC3	--RRAQDFVQWLM	HHHHHHHHHHHHH	HELICAL REGION
PLAA2	--ALWQFRSNTI-	--HHHHHHHHHBB	HELIX - SHEET

Abbreviations; GLC; bovine glucagon, PLAA2; porcine phospholipase A2 ; B = ß-sheet is predicted ; T = ß-turn is predicted ; H = α- helix is predicted ; - either : 1) no residue aligns to this position, eg. when the right hand neighbouring residue is an N-terminus, or 2) no secondary structure is predicted in the region of this residue. The amino acids are abbreviated using standard single letter code.

The proteins actin, glucagon and phospholipase A2 are amongst the few substrates of transglutaminase for which X-ray crystal structures are available. All of the sequences (which are aligned against the glutaminyl residue of interest) apart from GLC3 are reactive towards transglutaminase. The "PREDICT" computer programme consists of a bank of secondary structure predictive algorithms which are applied to amino acid sequences to give a weighted average prediction which has a higher overall accuracy than the individual methods alone. In the test cases above this programme was generally successful in its prediction of the secondary structure surrounding the known substrate sites in these molecules. Table 9.2 includes further predictions of protein substrates of transglutaminase. In some of these cases secondary structure has been implicated by physicochemical methods (although X-ray structures are not available). In several of these cases the PREDICT program corroborates these secondary structural investigations. The residue numbers of these sequences may be found in Table 10.1 and 10.2.

TABLE 9.2

The use of the program PREDICT to predict the secondary structure associated with transglutaminase substrates of unknown conformation

PROTEIN	PREDICTED STRUCTURE	PROTEIN SEQUENCE
PGK 1	BBBBBBBBBTTTTTT	GVTDKISQVSTGGGA
CAS 1	----------BBBB	DKIHPFAQTQSLVYP
CAS 2	T--TTBBBBBBBB--	QNIPPLTQTPVVVPP
CAS 3	-TTTTTT--------	PFPKYPVQPFTESQS
CAS 4	BBBBBB-TT------	QSVLSLSQSKVLPVP
CAS 5	--T-TTTT-T-----	EKAVPYPQRDMPIQA
CAS 6	-----T-T--T-T-T	SKVLPVPQKAVPYPQ
CAS 7	HHHHHH---TTT---	IQAFLLYQQPVLGPV
LIP 1	HHHHHHHHHHHHH--	MVSEFLKQAWFIENE
BLAC1	BBBTTHHHHHHHHH-	RLSFNPTQLEEQCHI
BLAC2	THHHHHHHHH-----	NPTQLEEQCHI----
INS A	-BBBBTT--T-----	GIVEQCCASVCS---
INS B	-TBBBBTTT-TT---	ASVCSLYQLENYCH-
ENDOR	----TTTTBBBBBBB	FMTSEKSQTPLVTLF
HB A	-TTTT----TTTTT-	DLSHGSAQVKGHGK-
COLLO	TTTTTTTTTTT-T--	GGCSHLGQSYADRDV
FIB 1	--TTTTT----TTTT	LTIGEGQQHHLGGAK

Abbreviations: PGK 1:- Phosphoglycerate kinase (yeast), CAS 1 through to CAS 7; bovine ß-casein, LIP 1:- Lipocortin 1 (human placenta),BLAC1:- ß-Lactoglobulin (bovine),BLAC2:-ß-Lactoglobulin (bovine),INS A:- Insulin A chain (Human),INS B:- Insulin B chain (Human),ENDOR:- ß-Endorphin, HB A:- Haemoglobin A chain (Human),COLLO:-Type III collagen (bovine), FIB 1:- Fibrin γ -chain C-terminal peptide (human). Amino acids are abbreviated to standard single letter code. Dashes either represent the absence of a residue or the absence of structure predicted about a residue.

The secondary structural predictions of a number of the protein substrates of transglutaminase tested are in general accord with what is known about the structures of these proteins from physicochemical studies. For example many of the sites in ß-casein are predicted to have little structure or have some turn potential. This protein is highly susceptible to trypsin and other proteinases and this is evidence for a loosely packed structure. The PGK structural prediction is similar to that of actin (see Table 9.1) with which it shares some sequence similarity. Perhaps this sequence within the Gln_{388} mutant resembles the fourteen residue loop surrounding the reactive Gln_{41} in actin. The ß-endorphin structure is predicted to have some sheet and turn structure. This structure has also been suggested by n.m.r. studies of the peptide (Licharge et al., 1987). Lipocortin I is predicted to be helical at the reactive glutaminyl site within the N-terminal section, and this is a similar situation to that seen in phospholipase A2 (see Table 9.1). It may be that transglutaminase modifies the C-terminal region of ß-lactoglobulin in a folded state which resembles some of the other helical sites in proteins shown in Figs 9.1 and 9.2.

eg. human lipocortin (see section 1.10.3. and 1.7.8.2.), and porcine phospholipase A2 (see section 1.7.8.2.). Substrate sites within these proteins were both predicted to be N-terminal helices. The local structure, and sequences of these α-helical substrates may contain special features which enhance their substrate effectiveness. This may be necessary to overcome steric effects associated with the packing of side chains in α-helices relative to sequences which more closely approach random coil conformations.

It is possible that a common structure is induced during catalysis by features at the active site of transglutaminase. Helical wheel representations (where amino acid sequences are modelled as α-helical conformations) do not indicate a common pattern when this is applied to the known substrate sequences of transglutaminase.

9.3. DATABASE SEARCHES FOR SEQUENCE MOTIFS IN TRANSGLUTAMINASE SUBSTRATES

9.3.1. INTRODUCTION

Two substrate sequences which superficially appear only slightly similar may be considered by a computer program to be the closest matches available within a database of 10,000 protein sequences. Hence the amino acid sequences of many of the known substrate sites within protein substrates of transglutaminase were compared against a computer database of 10,000 protein sequences to see if one substrate site could successfully predict another. In such an event it is

possible that significant sequence motifs might be highlighted within the current dataset. The general background to the use of computers in database searches for similar polypeptide sequences is described elsewhere, see section 2.2.8.4.. Peptide sequences of between 12 - 15 amino acids were used for each search, when not limited by constrains such as N- or C- termination of the polypeptide chain. The reactive glutaminyl residue was located centrally within the search sequence.

9.3.2. METHODS

Database searches were conducted using two different approaches. The first used the FASTA program of Pearson and Lipman (1987) which is incorporated in the "Profile search" facility of the GCG sequence analysis system. This program aligned sequences purely on the basis of overall similarity and resulted in the location of similar sequences to the search sequence. However in many cases the peptides which were identified as having overall similarity to the target sequences could not have substrate properties for transglutaminase due to the absence of a centrally placed glutaminyl residues within the sequence. The peptide sequences which resembled the search sequences were then searched manually to locate peptide sequences which contained glutamine residues at positions equivalent to the search sequence. **This work was done with the assistance of Dr. P. Cleat of the Biological Support group, Dept. of Computing, Edinburgh University.**

The second set of computer matches was done using the method described in section 2.2.8.4. This involved the matching of the glutaminyl residue of a peptide sequence against a protein sequence database, followed by automatic alignment of the rest of the sequence on the basis

of similarity of surrounding sequence. This work was done by Dr. A. Coulson, Department of Molecular Biology, University of Edinburgh.

When considering the significance of the similarity of matching a peptide sequence within a database against a search sequence, certain criteria should be considered eg.

1) If the glutamine within the matched sequence is known to not be exposed at the surface of the parent protein, then it is highly unlikely that it could act as a substrate for transglutaminase. In such cases a match is likely to be spurious and is of little interest.

2) If the matched sequence is derived from a highly exotic source of protein then it is most unlikely that the significance of the match may be experimentally tested, and thus for current purposes cannot be considered significant.

9.3.3. RESULTS AND DISCUSSION

9.3.3.1. GENERAL RESULTS

Both methods of computer matching located similar sequences to the parent sequence within other proteins, but due to differences in the statistical weighting associated with the search programmes some differences were seen between the lists of the best matched proteins. Thus the two methods may be regarded as complementary.

Database searches using the known modification sites for liver transglutaminase were often unsuccessful in finding matches with other reactive sequences. However search sequences were often matched against proteins which are known to be substrates for transglutaminase *in vitro*, but for which no substrate site(s) have been determined. It was not possible to say whether such matches were significant. These matches may be considered predictions of substrate sites, and will no doubt be useful in construction of theories concerning the specificity of transglutaminase as the location of reactive glutaminyl residues within these proteins are determined. Matches of potential significance included: myosin, apolipoprotein B-100, cartilage specific proteoglycan, troponin, tubulin, thrombospondin, and vitellogenin all of which have been reported to have substrate properties for transglutaminases.

In only a single case were two substrate sequences found to have high statistical similarity with each other. In this case the two substrate sites were located within the same molecule ie. ß-casein. This match is discussed in section 9.3.3.2.

9.3.3.2. TWO MATCHED SEQUENCES ARE LOCATED IN BOVINE ß-CASEIN

Glutaminyl residue Gln_{167} is known to be the best glutaminyl substrate for factor XIIIa within bovine ß-casein (Gorman and Folk, 1980). This site is also known to be a labelling site of the tissue enzyme in amidated, and succinylated ß-casein (Yan and Wold, 1984). Computer matching found high similarity [relative to the other sequences in the database] with another known labelling site in bovine ß-casein ie. sequence Gln_{72} through to Pro_{86}. This site is fully reactive towards transglutaminase in succinylated casein (Yan and Wold, 1984) and appears to be the major site of labelling of the tissue enzyme in native [non-succinylated, non-amidated] ß-casein (see chapter 7). The sequences of these two modification sites within ß-casein are aligned in Fig 9.1. The similarity appears to be greatest about the labelled glutamines and on the C-terminal side of the labelling sites.

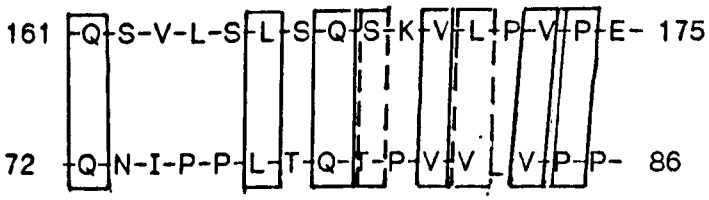


Fig 9.1. A COMPARISON OF THE MAJOR LABELLING SITES OF FACTOR XIIIa AND TISSUE TRANSGLUTAMINASE WITHIN BOVINE ß-CASEIN

Residues at identical positions are boxed with a solid line. Other similar motifs are boxed with hatched lines

Other known substrates of transglutaminase which were considered to have similarity with the sequence surrounding Gln_{167} included epidermal keratin (see section 1.10,5.), haemagglutinin (see section 14.2.1.), and myosin. The labelling sites are not known in each case but the haemagglutinin sequence maps to a surface loop within the X-ray structure of the protein (Wilson et al., 1981). Since labelling is decreased by limited tryptic digestion of the protein this loop may represent a genuine substrate site [see section Appendix I (iii)].

9.4. GENERAL CONCLUSIONS

Transglutaminase has been shown to have substrate properties towards many diverse amino acid sequences in proteins (see Chapter 1). It is therefore not surprising that the sequences associated with non-physiological substrates were not effective at predicting other known sites. Some of the matches may have identified reactive glutaminyl residues but it is presently impossible to predict *a priori* which sequences these might be. The substrate sequences associated with physiological substrates of the enzyme were no more effective in predicting further substrate sites than non-physiological substrates. This may merely reflect the presently small number of known modification sites within proteins, but more likely reflects the requirement for knowledge of the local secondary structure and the degree of exposure of glutaminyl residues of both the search sequence and the proteins to which these sequences are to be compared within the database. These factors are discussed in greater depth in chapter 10.

CHAPTER TEN: A PRELIMINARY SET OF RULES TO ACCOUNT FOR TRANSGLUTAMINASE SPECIFICITY

10.1. INTRODUCTION

The substrate sites of transglutaminase within proteins have been reviewed in chapter 1. In the absence of a clear consensus sequence motif surrounding reactive glutaminyl residues, conformational factors must be predominant in the determination of specificity of transglutaminase (see section 1.7.9.). Since some substrates (particularly those thought to be physiologically important *in vivo* eg. actin, phospholipase A2, and uteroglobulin appear to have discretely folded local structure (see section 1.7.8.2. and 1.7.8.3.) it is not appropriate to group these substrate sites with more conformationally flexible substrate sites when looking for consensus features.

In order to compare sequence motifs from similarly folded substrate sites, transglutaminase substrates have been divided into two broad structural groups, type A and type B. The distinction is made on the basis of available physical evidence and when this is not available on the basis of computer based secondary structural predictions, thus:

Type A substrates : either show little ordered structure (e.g. by n.m.r. or c.d.) or show low or indefinite electron density in X-ray diffraction studies, or are denatured, chemically modified proteins. They are not predicted to have high secondary structural potential by the computer programme PREDICT (see section 9.2.1. and 9.2.2.).

Type B substrates show distinct electron density (when an X-ray structure is available) or show strongly predicted secondary structure (e.g. lipocortin I) or are predicted to be similar to proteins of known structure e.g. the seminal vesicle protein SV-IV is predicted to be similar to uteroglobin which is a small compact protein (Metafora et al., 1987). The classification of the known modification sites of transglutaminase into type A and type B is shown in Tables 10.1a and b.

Since the importance of the distribution of sterically bulky (see section 1.7.9.3.) and charged residues (see section 1.7.9.4.) in the region surrounding substrate sites have emerged as likely determinants of the substrate properties of proteins, type A and type B substrates were inspected in order to see if any distinct patterns were evident (see section 10.2.).

10.2. COMPARISONS OF AMINO ACID SEQUENCE WITHIN TYPE A AND TYPE B SUBSTRATES OF TRANSGLUTAMINASE

10.2.1. INTRODUCTION

The percentage occupancy of each amino acid type at each position within a window, -5 residues +5 residues from the modified glutamine within type A and type B substrates was calculated. Substrate sequences of eleven or so residues are considered since this is comparable in length [35 Å, assuming 3.5 Å per amino acid residue (Schechter and Berger, 1967)] to the longest extended binding sites estimated for a number of proteinases (Fruton, 1975). The different amino acid types were grouped according to their chemical properties and the overall

TABLE 10.1 (a) TYPE A SUBSTRATES

SEQUENCES AROUND KNOWN SITES OF TRANSGLUTAMINASE-CATALYSED MODIFICATION OF PEPTIDES AND PROTEINS

Protein/peptide Type A	Sequence -5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	Reference
ACTH (Q11)	T	S	E	K	S	Q	T	P	L	V	T	a
Glucagon (Q3)				H	S	Q	G	T	F	T	S	b
Substance P (Q5)		R	P	K	P	Q	Q	F	F	G	L	c
Insulin												
(oxidised A chain)												
(Q5)		G	I	V	E	Q	C	C	A	S	V	b
(Q15)	V	C	S	L	Y	Q	L	E	N	Y	C	
(oxidised B chain)												
(Q4)			F	V	N	Q	H	L	C	G	S	
*Melittin (Q25)	I	K	R	K	R	Q	Q					d
β-casein (succinylated)												
(poor) (I) (Q54)	I	H	P	F	A	Q	T	Q	S	L	V	e
(II) (Q56)	P	F	A	Q	T	Q	S	L	V	Y	P	
(III) (Q79)	I	P	P	L	T	Q	T	P	V	V	V	
(IV) (Q167)	V	L	S	L	S	Q	S	(K)	V	L	P	
(V) (Q175)	V	L	P	V	P	Q	(K)	A	V	P	Y	
(poor) (VI) (Q182)	A	V	P	Y	P	Q	R	D	M	P	I	
(VII) (Q194)	A	F	L	L	Y	Q	Q	P	V	L	G	
Fibrin peptide												
(Q398)	T	I	G	E	G	Q	Q	H	H	L	G	f
Haemoglobin (heat denatured α-chain)												
(poor) (Q54)	S	H	G	S	A	Q	V	K	G	H	G	g
β-lactoglobulin (in the presence of dithiothreitol)												
(Q155)	S	F	N	P	T	Q	L	E	E	Q	C	h
(Q159)	T	Q	L	E	E	Q	C	H	I			

The positions -5 to +5 represent the positions of amino acid residues relative to reactive glutaminyl residues in the primary amino acid sequence.

*Melittin has been reported to have an α-helical conformation in water at high concentrations and hence rather than being classed as a type A substrate it is possible that under the conditions of labelling that melittin more closely resembles a type B substrate (see Table 19.1 b.). If this were the case then positive charge at position -1 in the sequence of type B substrates would not be considered a discouraging feature (see table 10.3).

TABLE 10.1 (b) TYPE B SUBSTRATES

PROTEIN		SEQUENCE											REFERENCE
Crystallin β-Bp	(Q7)	S	N	H	E	T	Q	A	G	K	P	Q	i
Crystallin β-B3	(Q3)				A	E	Q	H	S	T	P	E	i
Crystallin α-A3	(Q6)	E	T	Q	T	V	Q	Q	E	L	E	S	i
Lipocortin I	(Q18)	I	E	N	E	E	Q	E	Y	V	Q	T	j
Phospholipase A2	(Q4)			A	L	W	Q	F	R	S	N	I	k
Collagen III aminopropeptide	(Q14)	C	S	H	L	G	Q	S	Y	A	D	R	l
ACTIN	(Q41)	G	R	P	R	H	Q	G	V	M	V	G	m
SV-IV	(Q9)	K	E	K	Y	S	Q	S	E	E	V	V	n
	(Q86)	R	S	R	F	A	Q	D	V	L	N		n
Phosphoglycerate kinase (His388→Gln388 mutant)	(Q388)	T	D	K	I	S	Q	V	S	T	G	G	o
Glucagon	(Q20)	D	S	R	R	A	Q	D	F	V	Q	W	p

Sequences are shown in the one letter code, with the modified glutamine at position 0. Brackets around lysines indicate that the side chain has been succinylated. The term "poor" is used to denote that the modification of the designated glutamine is only partial. The distinction between Type I and Type II substrates is explained in the text.

References to table

a Pucci et al. (1988a)
b Folk and Cole (1965)
c Pucci et al. (1988b)
d Perez-Paya et al. (1991)
e Yan and Wold (1984)
f Gorman and Folk (1980)
g Pincus and Waelsch (1968)
h Coussons et al., (1992 in press)
i Pepinsky et al. (1989)
j Cordella-Miele et al. (1990)
k Bowness et al. (1987)
l Takashi (1988)
m Porta et al. (1989)
n Coussons et al. (1991)
o Berbers et al. (1983)

distribution of each class of side chain in the two classes of transglutaminase substrate compared. The results are shown in Tables 10.2 a and b, and discussed in sections 10.2.2. to 10.4. The results of this analysis have been combined in section 10.5. to form the basis of a preliminary set of rules which may define transglutaminase specificity. These rules have been applied to predict substrate sites in a number of model systems in chapter 11.

10.2.2. DISTRIBUTION OF UNCHARGED RESIDUES

Within type A substrates the large non polar and polar residues ie. tryptophan, methionine, isoleucine and phenylalanine (non-polar) are present in low abundance, although phenylalanine is present in slightly higher amounts. Those positions in which they have occupancy tend to be either N-terminal to the labelling site, or remote from the labelled glutamine (apart from phenylalanine which has been found at positions +2 and +3). Tyrosine [large polar (hydroxylated) side chain] is similarly distributed to the other residues with large side chains, with the exception that it has a relatively high occupancy as the N-terminal neighbour of the labelled glutamine (cf. small polar amino acid side chain distribution within substrate sites).

The smaller non-polar residues are present in much higher amounts than other amino acids within substrate sites and tend to have high occupancy in the region C-terminal to the labelling sites. This probably reflects a requirement for a hydrophobic component within substrates in order to enhance non-specific binding of substrates at the hydrophobic active site of transglutaminase. This is

particulary noticeable for Val, Leu, and Pro within positions +2 to +5 (the same region where charged residues are disfavoured). Alanine is not abundant in any region except position -1, where the only non-polar residues represented are Ala and Pro.

Serine and threonine (small polar side chains) are quite evenly distributed on either side of the labelling sites in all cases, with slightly higher occupancy in the region neighbouring the modified glutamine than elsewhere within substrate sites.

Glycine residues are predominantly located on the C-terminal side of labelling sites. This is in accord with a requirement for local flexibility in substrate sites. This phenomenon has been noted in a number of proteins, eg. the "hinged lid" of the triose isomerase loop (Joseph et al,. 1990), the "hinge" region of PGK [Watson et al., (1982), see chapter 6] and antibody molecules (Dangl et al., 1988). Since glycine residues lack side chains and are not generally present as a nearest neighbour to reactive glutaminyl residues, this may reflect a local steric requirement at this point which may be contributed by the adjacent residue. Yan and Wold (1984) have noted that the carboxamide group of glutamine should be able to approach the hydroxyl group of neighbouring Ser and Tyr residues using molecular modelling techniques. The potential of hydroxyl groups to increase the leaving group potential of ammonia from peptidyl glutamine residues could concievably contribute towards the stereochemical and chemical requirements of transglutaminase for glutaminyl-substrates at its active site (Wold, 1985).

The remaining uncharged polar residues ie. Asn, and Gln are present in low abundance at most positions. However glutamine has high occupancy of position +1 to the labelled glutamine in type A substrates. It may be significant that a number of the sequences shown in Table 10.1 a.b. possess a repeat sequence (i.e. -X-X- and -X-X-X-, where X represents a particular amino acid type).

Type B substrates have much the same distribution of amino acids about labelled sites as type A substrates, with a notable increase in the occupancy of glycine and alanine residues. This may reflect a requirement for increased local flexibility at certain positions within such substrates.

Although cysteine residues are present in a number of positions in insulin (type A substrate), they are oxidized to their sulphonic acid derivatives and are thus negatively charged. Since in all cases DTT was added to the labelling mixtures used it is unlikely that the -SH group of cysteine residues are important determinants of transglutaminase specificity. Interestingly cysteine groups are virtually absent within type B substrate sequences presumably reflecting their properties as flexibility restricting elements as disulphide bridges.

10.2.3. DISTRIBUTION OF CHARGED RESIDUES WITHIN SUBSTRATE SITES

10.2.3.1. POSITIVELY CHARGED RESIDUES

Since some of the lysine residues within the substrate sites of transglutaminase were succinylated the charge on these residues was reversed and hence they are classed as negatively charged residues. Histidine can become positively charged below a pH of 6.0 but since all the labelled sequences were determined from material which was prepared at pH 7.0 or above this residue should probably be classed as polar. The amino terminus of polypeptide chains is expected to carry a positive charge at this pH.

Within a total of eighteen type A sequences, seven positively charged residues (77 %) were located on the N-terminal side of the labelled glutamine (this includes N-terminal residues), and two (23 %) were located on the C-terminal side. Only eleven search sequences were available for type B substrates but the overall distribution of positive charge was very different to that of type A substrates ie. 30 % N-terminal, and 70 % C- terminal to the modified sites.

It has been noted elsewhere (see section 1.7.) that positive charge is often absent in transglutaminase substrates and that this therefore may reduce the substrate properties of proteins. This generalization appears to be particularly true for type A substrates. Only two sequences contained positively charged amino acid side chains at position +1 relative to the modified Gln residue (both in β-casein), and of these sites Gln_{175} was succinylated. The second such site (Gln_{182}) was reported to be a poor substrate relative to others labelled (60 % modified as opposed to 100 % modification. The presence of positively charged residues at position +2 may also reduce substrate properties, since Lys_{169} of β-casein had been succinylated prior to the enzymic modification of

Gln_{167} by transglutaminase (Yan and Wold, 1984) and if labelled represents only a minor site of modification in non-succinylated ß-casein (see chapter 7). Glutaminyl residue Gln_{54} in heat denatured haemoglobin was reported to be a poor substrate, though succinylation of Lys_{56} appeared to improve its substrate properties (Pincus and Waelsh, 1968). Examination of Table 10.3 indicates that positive charge may be disfavoured over a region of five or more residues along the C-terminal side of the polypeptide chain adjacent to a reactive glutaminyl residue in type A substrates. Positively charged residues were not observed in positions: -1, +1, and +4 in type B substrates. These results are summarized in table 10.3..

10.2.3.2. NEGATIVELY CHARGED RESIDUES

At pH 7.0 and above only the carboxyl terminal residue and the side chains of aspartic acid and glutamic acid are expected to be negatively charged in a typical protein. The sulphonic acid derivatives of cysteine side-chains are expected to resemble aspartic acid residues in their chemistry and therefore are considered in this section (see section 10.2.2).

Negatively charged residues were present in higher abundance within type A substrates than positively charged residues. Four (25 %) were N-terminal to the labelling site, and twelve (75 %) were C-terminal. Type B substrates contained seventeen negatively charged groups, nine are in the C-terminal region relative to the labelled site (56 %). Eight (44 %), are in the C-terminal region. The overall distribution of positions where a negatively charged residue is observed to be permissible differs between type A substrates and type B substrates. In type A

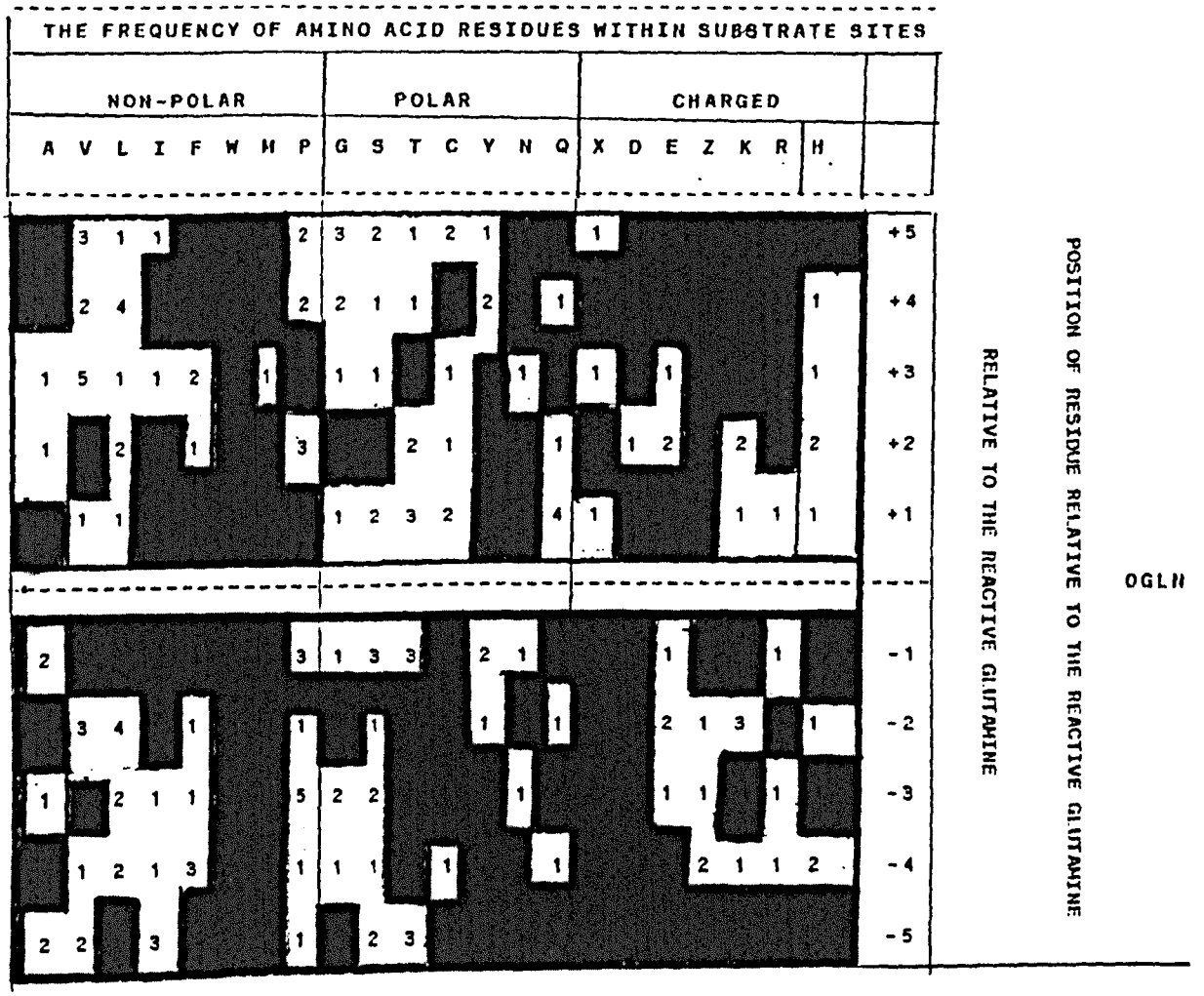
substrates there appear to be few positions where negative charge may be unfavourable to substrate properties ie. -5 and +4. In type B substrates negative charge was absent at a number of positions ie. -3 and +4, These results are summarized in Table 10.3..

10.3. A COMPARISON OF TYPE A AND TYPE B SUBSTRATES

The amino acids which comprise the substrate sites of transglutaminase have been classified according to their general chemistry and the frequency at which they are found at a given position relative to reactive glutaminyl residues is shown in Tables 10.2 a and b. The positions where a particular amino acid has not been identified are shaded. The overall distribution of uncharged amino acid residues within type B substrates is similar to that of the type A group, although generally less restrictive in terms of which residues may be allowed in certain positions. This may reflect the ability of more highly folded substrates to position otherwise unfavourable charged or sterically bulky groups away from important specificity determining regions within the active site of transglutaminases. Interestingly alanine and glycine residues are featured more regularly on both sides of the modified sites in type B substrates than in type A substrates. This may reflect a requirement in some type B substrates for local flexibility in order to overcome steric restrictions at the active site of transglutaminase. The distribution of charged residues is considerably different in type A and type B substrates and this may be an important determinant of the specificity of transglutaminase (see section 10.4).

TABLE 10.2 (a) TYPE A SUBSTRATES

The distribution of amino acid residues surrounding reactive glutaminyl residues in type A substrates



* The number signifies the position of a particular amino acid relative to a centrally located glutaminyl residue (position 0) within a sequence. Positive values signify residues C-terminal to the reactive glutaminyl residue and negative values represent residues located N-terminally to this residue.

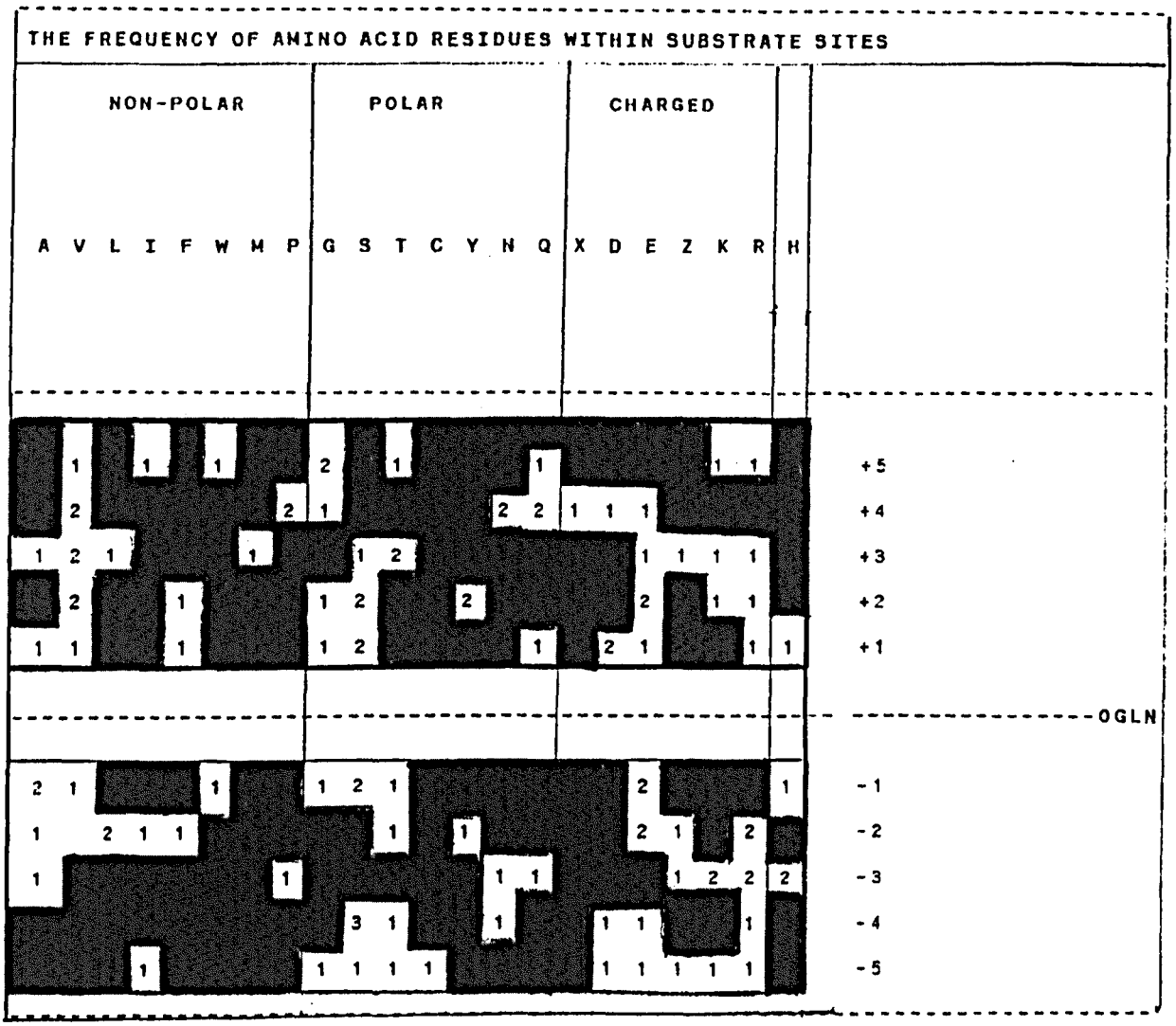
The amino acid residue frequency at a position within the total number of substrate sites known for a specific conformational class of substrate is indicated below the amino acid residue, which is indicated by single letter code. X in this case corresponds to a C-terminal amino acid, and Z corresponds to an amino terminal residue. These will carry a negative and a positive charge respectively under the conditions used for labelling proteins.

The amino acids are grouped according to the three broad classes of : Charged, polar and non-polar for ease of interpretation of trends in the sequences.

The same format is applied to type B substrates in table 10.2 (b) over page.

TABLE 10.2 (b) TYPE B SUBSTRATES

The distribution of amino acid residues surrounding reactive glutaminyl residues in type B substrates



Shaded areas indicate that no amino acid type specified in the heading is located at this position in any of the substrate sequences determined to date.

10.4. A CLASSIFICATION SYSTEM FOR TRANSGLUTAMINASE SUBSTRATES

10.4.1.INTRODUCTION

Since it has proved difficult to devise a theory which will predict the substrate properties of a given protein (X) on the basis of the presence of features within exposed regions of proteins, an alternative approach for understanding the specificity of features transglutaminase is considered. This is based on searching for features within proteins which apparently discourage modification of a particular glutamine side chain. From inspection of the distribution of charged residues within transglutaminase substrates (see section 10.2.3.) it appears that their presence/absence may fulfil the required criteria in the construction of such a predictive theory.

If, for example, a positively charged side chain appears at a certain position (say 4 residues on the N-terminal side) in any substrate, then it is concluded that such a residue in this position is not a discouraging feature. However if no positively charged side-chain appears at a certain position eg. four residues to the C-terminal side of reactive glutaminyl residues in all examples of that structural class then such charge at this position is classed as a discouraging feature. The various "discouraging features" are summarised in the set of "rules" shown in Table 10.3. Overall it appears that positively and negatively charged side chains to the C-terminal side of the glutamine side chain could play an important role in determining transglutaminase specificity. Gorman and Folk (1984) had obtained some evidence for this type of effect in their studies on model peptide substrates for the enzyme.

TABLE 10.3

CHARGED SIDE CHAINS WHICH ACT AS DISCOURAGING FEATURES IN THE TRANSGLUTAMINASE-CATALYSED REACTION

Position	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Positive charges											
Type A substrates	*					Q	(*)	(*)	*	*	*
Type B substrates					(*)	Q	*			*	
Negative charges											
Type A substrates	*					Q				*	
Type B substrates			*			Q					(*)

In this Table the presence of an asterisk at a particular position relative to the glutamine (at position 0) indicates that the given charged side chain would act as a discouraging feature. Such side chains are absent from the known substrates for transglutaminase listed in Table 1. The brackets indicate that these designations are less definite, since they refer to either relatively poor substrates (succinylated β-casein and the heat-denatured α-chain of haemoglobin) or an uncertainty in the sequence (crystallin βB3).

The distinction between Type A and Type B substrates is explained in the text. Since the majority of the modification reactions are carried out at a pH of 7.5 or above, histidine side chains are not included in the charged side chain category.

The rules predict that a (type B) substrate will not be reactive if they bind at the active site of transglutaminase in an extended conformation, since this conversion to a type A conformation would exclude proteins such as phospholipase A2 from being reactive. Conversely a type A substrate could be converted to a type B conformation and retain its reactive properties. A peptide with a positive charge at position +4 would not be predicted to be a substrate in either a type A or type B conformation.

10.5. A SUMMARY OF FACTORS WHICH DETERMINE TRANSGLUTAMINASE SPECIFICITY

On the basis of the available information concerning the protein and peptide substrates of guinea pig liver transglutaminase some features which apparently determine the specificity of transglutaminase are summarized below.

1) Peptides as small as two amino acids in length may act as (poor) substrates of transglutaminase. In order for proteins of up to nine residues in length to be efficient substrates for transglutaminase:

a) it will contain a glutaminyl residue. Asparagine (which has one less methylene group in its side chain) has no substrate properties.

b) it should be blocked at the N-terminus by a group such as the carboxybenzoyl group in CBZ-Gln-Gly (see Fig 2.3).

c) preferably be modified at the C-terminus eg. by esterification to an ethyl ester derivative.

d) obey the charge rules for type A (highly flexible) substrates.

e) not be cyclic in conformation.

2) Larger peptide substrates. The criteria for substrate properties is similar to smaller peptides. In addition they must:

a) not contain either a N- or C- terminal glutaminyl residue.

b) not fold into a conformation in which the glutamine residue is sterically blocked by neighbouring residues

c) not contain charged residues in positions which contravene the charge rules presented in section 10.4.. In some cases peptides of this size may have considerable secondary structure. In these cases it is important to class the peptide correctly as either type A or type B.

3) The criteria for larger peptides and proteins is broadly that described for peptides exceeding nine residues. However additional points should be considered if a protein has a globular conformation:

a) Glutamine residues must be sufficiently exposed in order to act as transglutaminase substrates. The lower limit is not yet known but appears to be between 50 and 70 % according to the criteria of Chothia (1975).

b) There is a requirement for high flexibility relative to the "framework" structure of a protein substrate in the region surrounding the glutaminyl site. Hence large surface loops and N- and C- termini are favoured relative to other structures.

c) Glutaminyl residues which are restricted in motion by the presence of disulphide bridges in a region of approximately -5 to +5 residues relative to the glutamine will have reduced substrate properties. This criteria clearly resembles that of rule 1) e. which relates to smaller peptide substrates.

4) The presence of amino acid side chains with negative charges or highly electronegative atoms (particularly hydroxylated amino acids such as threonine and serine) appears to improve substrate properties of peptide materials. Their high occupancy at positions neighbouring reactive glutaminyl residues may reflect more than their tendency to be present at high concentrations in flexible regions of proteins ie. they may directly improve the substrate properties of glutamine residues via specific polarizing effects.

5) The presence of certain uncharged amino acids may reduce substrate properties eg. cysteine, isoleucine and tryptophan. This observation is tentative and requires further investigation.

Factors 3), 4) and 5) may be considered to be linked in that the observed tendency of polar and charged amino acid side chains to occur at neighbouring positions to substrate glutamines may reflect the exposure of this part of the polypeptide chain to aqueous solvent and hence increase the accessibility of these glutaminyl residues relative to those which are not substrates for the enzyme. Sites which do not have substrate properties often are surrounded by non-polar, hydrophobic amino acids within the primary amino acid sequence. In addition hydroxylated amino acids (eg. serine, threonine, and tyrosine), could play a role in the polarization of the carbonyl group of the amide side chain of reactive glutaminyl residues through the formation of hydrogen bonds. This would then encourage attack by the active site Cys_{275} of transglutaminase (see section 1.5.) to form the thioester intermediate.

10.6. A PRELIMINARY ESTIMATE OF THE FREQUENCY OF OCCURRENCE OF REACTIVE GLUTAMINYL RESIDUES WITHIN GLOBULAR PROTEINS

A set of rules would be useful if it could predict the likelihood of a transglutaminase catalysed modification of a glutaminyl side chain in a given sequence. Since the list of known sites of modification is small a method of database searching has been developed, if only to attempt to define the scale of the problem more exactly. Glutamine occurs about 4 times per 100 residues (some 266873 times out of 6525000 residues) in the SWISSPROT database [Release 17 (see Bairoch and Boeckman (1991)]. No convenient reliable method is available for predicting how many of these residues are likely to be accessible. However from the Brookhaven protein structure database, the BIPED procedure (Akrigg et al., 1988) reveals that

8.2 % (229 out of 2778) glutaminyl residues have relative accessibilities (Chothia 1975) greater than 0.66 in the fully extended conformation. Thus about 8.2 % of glutaminyl residues in globular proteins are putative substrates for transglutaminase. This would be about 22000 residues in the current SWISSPROT database. The choice of 0.66 is somewhat arbitrary but from an analysis of the accessibility of unreactive and substrate sites in polypeptides such as glucagon (see section 11.3.3.) this figure may be considered reasonable. It should be noted that if a relative accessibility of greater than 0.85 is chosen then the number of potentially reactive glutaminyl residues in the Brookehaven database drops to 34 out of 2778 or 1.2 %. This would amount to 3260 side chains in the SWISSPROT database.

Next using the charge rules developed in this thesis for those features which discourage the reaction this number is further reduced. A negative charge has as frequency of about 11.5 % while a positive charge has a frequency of 11.1 %. Assuming that the occurrence of side chains is random, the probability that at least one of the six discouraging features for <u>type A</u> substrates will not occur in a glutamine containing peptide is about 50 % (0.89^6) . Thus using a relative accessibility of greater than 0.66 there must still be on average one in every 24 glutamines which (about 4.1% or 11000 in the database) is in a position to be modified by transglutaminase. Even applying a more rigid accessibility criteria of 0.85, this number is still about 3200. This is only a very rough estimate since it is probable that;

1) the rules are not yet sufficiently developed to correctly predict substrate sites in every case,

2) that X-ray structures within the database may not always reflect the flexibility of potential substrate sites in free solution.

3) The assignment of secondary structure to the protein sequences within the SWISSPROT database will contain many mistakes since structure was assigned on the basis of the results of computer based algorithms used for predicting secondary structure. Moreover, the concentration of charged residues about glutaminyl residues on the surface of globular proteins is likely to be higher than that surrounding residues which are in internal positions and not a random distribution as suggested in the calculation.

This approach corroborates the general finding that few glutaminyl residues in proteins are likely to be highly reactive towards transglutaminase. Those globular proteins that do have substrate properties will rarely contain more than one or two glutaminyl residues which may be modified by transglutaminase when the substrate is in its fully folded state.

CHAPTER ELEVEN: THE APPLICATION OF THE RULES GOVERNING THE SPECIFICITY OF TRANSGLUTAMINASE TO SOME MODEL SYSTEMS

11.1 INTRODUCTION

In order for the rules presented in chapter 10 to be useful, they must be able to successfully predict the sites of modification of transglutaminase within protein substrates. Since the substrate sites determined as a part of this thesis ie. Gln_{79} of bovine ß-casein, Gln_{155}, and Gln_{159} of ß-lactoglobulin, Gln_{388} of PGK (mutant protein), and Gln_{308} of porcine pepsin were not included during the original formulation of the rules, the rules were applied to these substrates, and the results are discussed in sections 11.2. to 11.4..

The substrate characteristics of a number of other protein substrates which are not so well structurally characterized are discussed in relation to the rules in Appendix I of this thesis. The rules were also applied to a number of protein substrates of transglutaminase which were used to devise the rules. In these cases it was particularly important for the rules to correctly predict those glutaminyl residues which although exposed do <u>not</u> act as substrates.

11.2. APPLICATION OF THE RULES TO PROTEIN SUBSTRATES DETERMINED AS PART OF THIS PROJECT

11.2.1. BOVINE β-LACTOGLOBULIN

From the X-ray structure (Papiz et al., 1986) it appears that the side chains of Gln_{13}, Gln_{35}, Gln_{68} and Gln_{120} are not significantly exposed to the solvent (see Fig 5.7.). Of the remaining glutamine side chains, Gln_5 is classified as type A (indefinite electron density) and Gln_{59} and Gln_{115} as type B. Little is known about the environments of Gln_{155} and Gln_{159}, because these are in a poorly defined C-terminal loop. However, in the presence of dithiothreitol, the disulphide bond (Cys_{66} - Cys_{160}) is presumably broken and the C-terminal loop would become flexible. Application of the charge rules shows that Gln_5 is not a substrate (lysine at position +3), nor are Gln_{59} (lysine at +1) or Gln_{115} (glutamic acid residue at position -3). Residues, Gln_{155} and Gln_{159} however are predicted to be allowed as substrates; this correlates with the data presented in section 6.3.. Glutaminyl residue Gln_{115} which has previously been suggested to be a third reactive site for transglutaminase (at pH 9) is the only other residue within the primary amino acid sequence not to be excluded as a substrate if β-lactoglobulin is considered to be substantially denatured at high pH and hence to contain all type A substrate sites.

11.2.2. PHOSPHOGLYCERATE KINASE

In chapter 6 it was shown that wild type yeast phosphoglycerate kinase is not a substrate for transglutaminase. Of the eight glutamine side chains in the enzyme, six are not exposed to the solvent, but Gln_{135} and Gln_{145} are in the "nose"

region of the N-terminal domain which is highly flexible on the basis of X-ray studies (Watson et al., 1982). Applying the rules for Type A substrates, both Gln_{135} and Gln_{145} would not be allowed as substrates. Gln_{135} has lysine at +1 and arginine at +3; Gln_{145} has lysine residues at positions +1, and +3. Hence, the lack of reaction can be accounted for. Gln_{37} is moderately exposed in the crystal structure of PGK. This residue is the first within a helical region of eight residues and hence must be a type B substrate. The lack of reactivity of this residue may be accounted for by the positive charge carried by Arg_{38} i.e. at position +1 which is disfavoured by the rules.

11.2.3. BOVINE β-CASEIN

The tertiary structure of β-casein is not known. However from the low stoichiometry of labelling it appears that under the conditions used for modification of the protein by transglutaminase that β-casein is substantially folded. This was also concluded from c.d. studies of β-casein (see section 7.2.1.).

The sequence surrounding the modification site which was determined [ie. Gln_{79}], did not contravene the charge rules. It could be argued that this result is not surprising given that this same sequence was used in the construction of the rules (see section 10.4.). Such a criticism is not so simply justified, since β-casein was not denatured under the conditions used (see chapter 7), whereas the modification site was originally determined in succinylated β-casein which was considered to be largely unfolded. Thus despite the constraints imposed on this site in the folded form of β-casein relative to the largely denatured succinylated form, Gln_{79} retained substrate properties.

11.3. APPLICATION OF THE RULES TO OTHER PUBLISHED PROTEIN SUBSTRATES OF TRANSGLUTAMINASE

11.3.1. RABBIT MUSCLE ACTIN

The X-ray structure of actin (see Fig 1.11.) shows that there is an exposed loop extending from Arg_{38} to Tyr_{54} (Kabsch et al., 1990). Within this sequence Gln_{41} can be extensively modified, whereas Gln_{50} is only slightly modified (Takashi, 1988). Application of the rules to this loop shows that Gln_{41} does not possess any neighbouring discouraging features, whereas Gln_{50} has a lysine at the +1 position, and so would be disallowed. This argument would apply if the loop were either flexible or folded (i.e if the glutamines were either type A or type B substrates). Although the loop possesses definite electron density in the actin-DNAase complex (Kabsch et al., 1990), it is not clear whether this would also be the case for actin itself, since the loop is involved in interactions with the DNAase.

11.3.2. BOVINE β-Bp CRYSTALLIN

There are three regions of the β-Bp-crystallin molecule where there is low electron density in the X-ray structure (Bax et al., 1990). These sites may be conformationally flexible and hence glutaminyl residues within these regions may be considered to be potential substrate sites for transglutaminase. These regions are located in the N-terminal 15 amino acid arm (which includes Gln_7 and Gln_{12}), in the connecting peptide (Ile_{99} - Lys_{107}, which contains Gln_{104}), and in the C-terminal region (which includes Gln_{193} and Gln_{196}) (see Fig 1.13.). The electron density data within this region are inconclusive, so the glutamine residues within

this region may be classified as either type A or type B. However, in either case both Gln_{193} and Gln_{196} are ruled out as substrates since Arg_{197} represents a discouraging feature at the +4 or +1 position respectively. In the connecting peptide, Gln104 would not be modified since according to the rules for type A substrates, the lysine at +3 would represent a discouraging feature. Reaction of Gln_{104} may also be ruled out by the hydrogen bond formation which is thought to occur to the side chain of Asp_{102} (Bax et al., 1990). The N-terminal region is more complex to analyze. Although the electron density data are inconclusive, secondary structure predictions and comparisons with other related structures (Wistow et al., 1981) suggest that Gln_7 should be at the C-terminal end of a short helix which is followed by a reverse turn and a coil which includes Gln_{12}. It is possible that Gln_7 is a type B substrate, and although there is a lysine at the +3 position, this would not be a discouraging feature. On the other hand, Gln_{12} would probably be in a flexible region (type A substrate) and thus the presence of a lysine at +5 would be a discouraging feature. These predictions are consistent with the observed modification data, which show that Gln_7 is the only site of modification (Berbers et al., 1983).

11.3.3. BOVINE GLUCAGON

There are three glutamine side chains in glucagon (Gln_3, Gln_{20} and Gln_{24}). From the X-ray structure (Sasaki et al., 1975; Blundell and Wood, 1982), Gln_{24} is rendered inaccessible [accessibility = 0.49 (Chothia, 1975)] by the side chains of Asp_{21} and Leu_{26}, see Fig 1.7. and hence is expected to have poor substrate properties (Gln_{24} would in any event be ruled out according to the rules by the

presence of the aspartate at the -3 position). Glutaminyl residue Gln_3 is in a flexible region and hence classified as a type A substrate (accessibility 0.84), whereas Gln_{20} is in a folded region (type B substrate) but with the side chain exposed (accessibility = 0.74). Application of the charge rules shows that Gln_3 and Gln_{20} should both be substrates, since there are no discouraging features. This is observed to be the case; transglutaminase catalysed modification occurs at Gln_3 and Gln_{20}, but not at Gln_{24} (Folk and Cole, 1965).

11.3.4. BOVINE TYPE III COLLAGEN

The aminopropeptide of calf collagen type III (Col 1-3)(III) contains a glutamine at position 14. There are no X-ray structural data for this peptide, but in view of the cysteine at position 9 being involved in a disulphide bond, and taking into account structure predictions (Horlein et al., 1979), it is reasonable to assume that Gln_{14} is in a region of folded structure (type B substrate). Using the charge rules, there are no discouraging features and it has been observed that Gln_{14} can act as a substrate for transglutaminase (Bowness et al., 1987). However, when the disulphide bonds in the protein are broken, it appears that Gln_{14} is no longer a substrate (Bowness et al., 1987). This behaviour can be understood if this glutamine is now considered to be in a flexible region (type A substrate) in which case the aspartate at +4 and the arginine at +5 act as discouraging features. It may be significant that the equivalent peptide from sheep has an arginine instead of aspartate at the +4 position relative to glutamine and even in the non-reduced, folded (type B) form of the peptide the glutamine cannot act as a substrate. This

is consistent with the charge rules which show that this arginine would act as a discouraging feature.

11.4. APPLICATION OF THE RULES TO GLOBULAR PROTEINS WHICH LACK SUBSTRATE PROPERTIES

The rules appear particular successful in the prediction of unreactive sites within proteins which also contain reactive sites (see section 1.7.8.4.). A lack of reactive sites is correctly predicted in the following proteins ; phosphoglycerate mutase, hen egg lysozyme, pancreatic and ribonuclease A. The structures of these proteins are discussed in relation to their substrate properties in section 4.3.1..

11.5. LESS WELL CHARACTERIZED SUBSTRATES OF TRANSGLUTAMINASE

The examples used to show the effectiveness of the rules in sections 11.2 to 11.4. were chosen principally because they are amongst the best structurally characterized substrates amongst those proteins, tested as substrates. The rules also correctly predict that there are two substrate sites in rabbit muscle aldolase, and two sites in bovine liver catalase. The structures of these proteins are shown and discussed with regard to their substrate properties in section 4.4..

11.6. CONCLUSIONS

By dismissing glutaminyl residues within proteins which have features which apparently contribute to the lack of reactivity of glutaminyl residues towards transglutaminase, it seems possible to predict which of the remaining glutaminyl residues will be reactive on the basis of their exposure. The success of this approach will be further tested as more substrate sites are identified in globular proteins.

Unreactive glutaminyl residues are often highly exposed and thus the importance of the distribution of local charged residues about substrate sites seems to provide a suitable basis for understanding why some exposed residues are reactive and others are unreactive. Unfortunately no structure is presently available for transglutaminase so the contribution of charged side-chains at the active site of transglutaminase to its specificity cannot be assessed. A possible chemical basis underlying the rules is discussed further in chapter 12.

CHAPTER TWELVE: MODELLING THE ACTIVE SITE OF TRANSGLUTAMINASE ON THAT OF PAPAIN

12.1. INTRODUCTION

Since structure and function are often inextricably linked in biological systems one of the best approaches to understanding the basis of transglutaminase specificity would be to examine the conformation and chemical features associated with the active site of transglutaminase or a related enzyme. An X-ray structure is currently not available for transglutaminase, but there is one available for papain, with which transglutaminase shares some sequence similarity (see section 1.4.). Both enzyme families catalyse esterase reactions and are postulated to have extended active sites.

The secondary structure of the amino acid sequence located at the active site of transglutaminases was predicted using computer based methods (see chapter 9) and this was compared using computer graphics to the folded structure (Drenth, 1968) of papain (see Fig 12.1.). The similarity of amino acid sequence between the active site of transglutaminase and the active site of papain (Ikura et al., 1988) suggested the possibility that amino acid side-chains common to both proteins might have similar roles in the catalytic mechanism of both proteins, and/or form similar structures in both proteins.

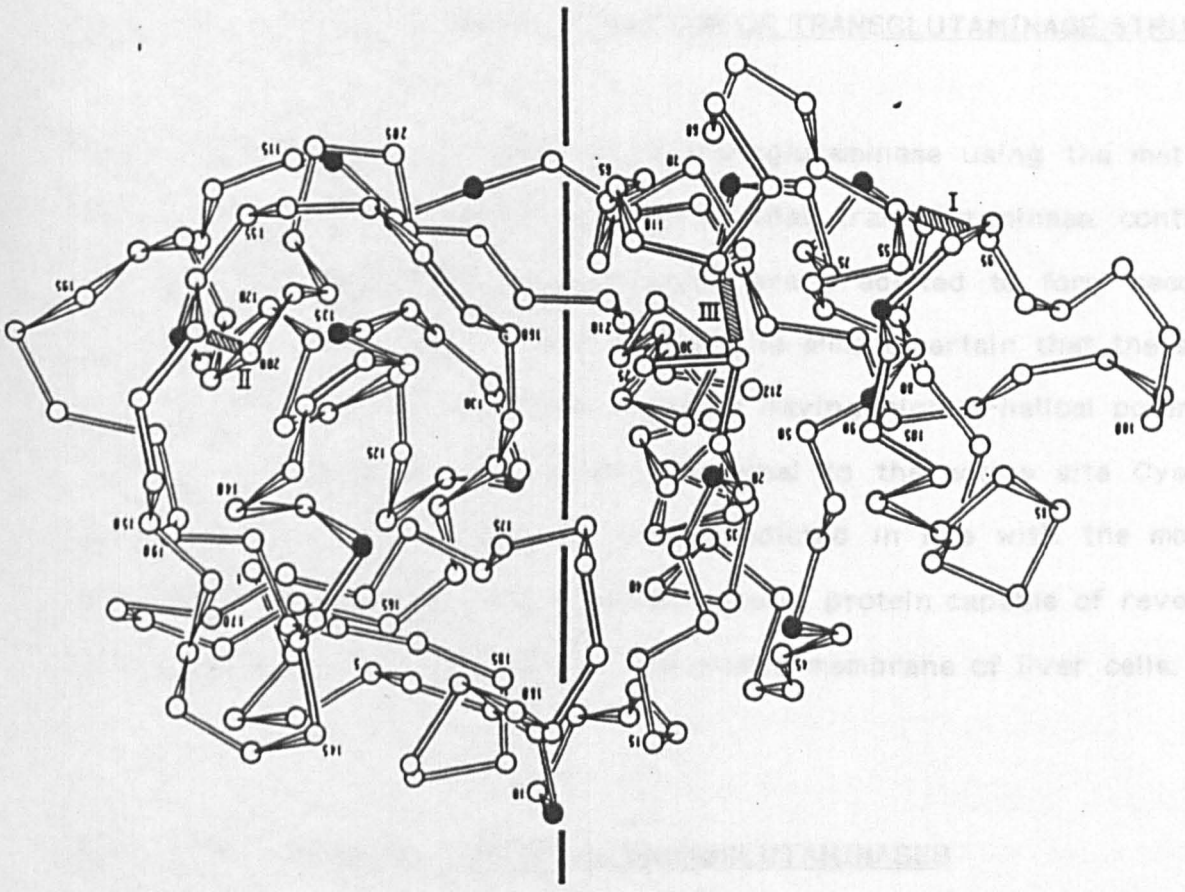


Fig 12.1. THE TERTIARY STRUCTURE OF PAPAIN

The α-carbon structure of papain from papaya is taken from Drenth et al., (1968). The protein is divided into two lobes and the active site is located in the central cleft. A number of residues are implicated as being important in the catalytic mechanism of the enzyme and a model of the orientation of these groups is shown in Fig 12.5.. It is possible that papain shares some of its structural features with transglutaminase (see text).

12.2. SECONDARY STRUCTURAL PREDICTION OF TRANSGLUTAMINASE STRUCTURE

A secondary structural prediction of transglutaminase using the method of Chou and Fasman (1974 a,b) suggested that transglutaminase contains a significant proportion of residues which are predicted to form secondary structure ie. ß-sheet and α-helix. Hence it is almost certain that the enzyme contains some globular structure. A region having high α-helical potential is particularly marked in the region C-terminal to the active site Cys_{275}. No transmembrane spanning regions were predicted in line with the model of transglutyaminase being an extrinsically bound protein capable of reversably binding to the cytoplasmic face of the plasma membrane of liver cells.

12.2.1. THE CYSTEINE $_{275}$ MOTIF IN TRANSGLUTAMINASES

The region surrounding Cys_{275} in liver transglutaminase is highly conserved in all known transglutaminases and resembles sequences present at the active sites of cysteine proteinases. Secondary structural predictions of liver transglutaminase were conducted using the cDNA derived sequence of the liver enzyme, (Ikura et al., 1988). Hydropathy analysis using the methods of Goldman et al. (198*), and Kyte and Doolittle (1982) confirmed the observation of Ikura et al., (1988) that the active site Cys_{275} is predicted to lie midway between a highly hydrophilic stretch of amino acids which extend N-terminally for sixteen amino acids to Arg_{259} before being broken by the hydrophobic Leu_{258}, and a highly hydrophobic region which extends for twelve residues before being broken by Arg_{287}. Cysteine Cys_{275} is predicted to lie in a flexible antigenic region which extends from Ser_{258} to Trp_{277} within a region

predicted to be composed of either ß-turns or ß-sheet depending on the method used. Residues Ala_{230} to Arg_{235} are predicted to form a relatively inflexible region with little antigenicity. This would be compatible with the hydrophobic pocket which is thought to lie in the vicinity of the active site cysteine (Folk and Gross 1971). The sequences which surround the active site cysteine residue of a number of transglutaminases are aligned against each other and papain in Fig 12.2.

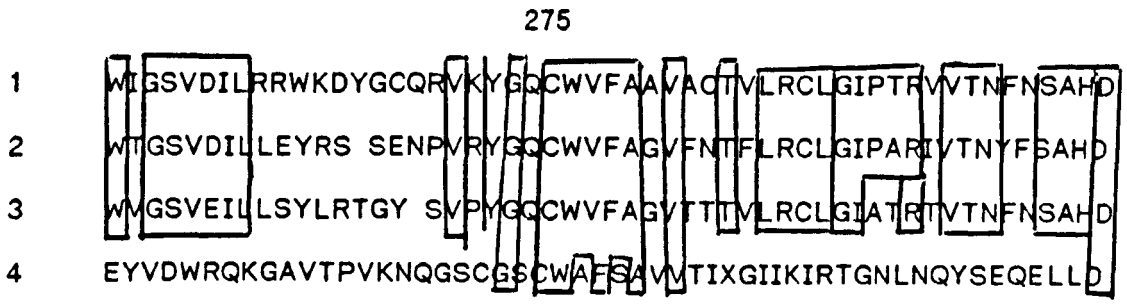


Fig 12.2. ALIGNMENT OF THE ACTIVE SITE SEQUENCES OF TWO TRANSGLUTAMINASES AND PAPAIN AGAINST CYS_{275} OF GP LIVER TGASE [1 = GP LIVER. 2 = FACTOR XIIIa. 3 = KERATINOCYTE (HUMAN)].[4 = PAPAIN]. Gaps have been introduced into the guinea pig liver sequence in order to maximize the similarity between sequences

The three transglutaminases are most similar to each other than to papain, although several residues in the region of the active site cysteine are conserved in all cases eg. Glycine at position -2, Tryptophan at +1, Phenylalanine at +3, and valine at +5. The transglutaminase with the highest similarity to papain is the keratinocyte transglutaminase. However the similarity between the liver, keratinocyte, and the plasma enzyme is sufficiently strong for the folding of the polypeptide chain to be

predicted to be similar. The charge profiles of the active site regions of transglutaminases (1-3) are shown to be similar in the region C-terminal to the active site Cys_{275}, but somewhat different in the region N-terminal to this residue. If these N-terminal residues contribute their side chains to the active site of these proteins this could provide the basis for differences in the observed specificity between these enzymes, since the distribution of charged residues within substrate sequences could influence electrostatic attraction or repulsion during the binding of substrate to an extended substrate site.

12.3. A COMPARISON WITH THE TERTIARY STRUCTURE OF PAPAIN

From the X-ray structure of papain, the side chains of Gln_{19}, Ser_{24}, Cys_{25} and the α-carbon of Gly_{23} project into the active site groove. Various aspects of the structure of papain are reviewed by Drenth et al. (1971). The active site Cys_{25} lies at C-terminus of a loop structure which is partly external and at the beginning of a long α-helix (extending from residue Lys_{17} to Arg_{41}, approximately 34 Å in length) which forms the lip of one domain within the active site cleft of the protein, (see Fig 12.2.). Although it is highly speculative to suggest that a similar region of α-helix is located in regions C-terminal to the reactive cysteine side-chain in both the transglutaminases and the cysteine proteinases it is not unusual for discrete structural motifs to be associated with diverse proteins catalysing reactions which have common mechanisms. The similarity in sequence (and possibly folding) may be a consequence of either:

1) divergence of both families of enzyme from a common ancestral protein.

2) convergent evolution producing similar structures (see section 1.4.).

12.4. THE HIS_{20} MOTIF IN TRANSGLUTAMINASES

A second conserved region within the primary sequence of transglutaminases extends from Arg_{12} to Thr_{22} in guinea pig liver transglutaminase. Like the Cys motif (see section 12.3.) this region may have a homologue in papain. Within papain, both the main chain and R-groups of His_{159} and Ala_{160} are located within 4 - 5 Å of Cys_{25} and this residue is considered the best positioned of all nearby chemical groups to participate in the mechanism of reaction (Drenth et al., 1968). Related sequences in transglutaminase and cysteine proteinases are aligned in Fig 12.3..

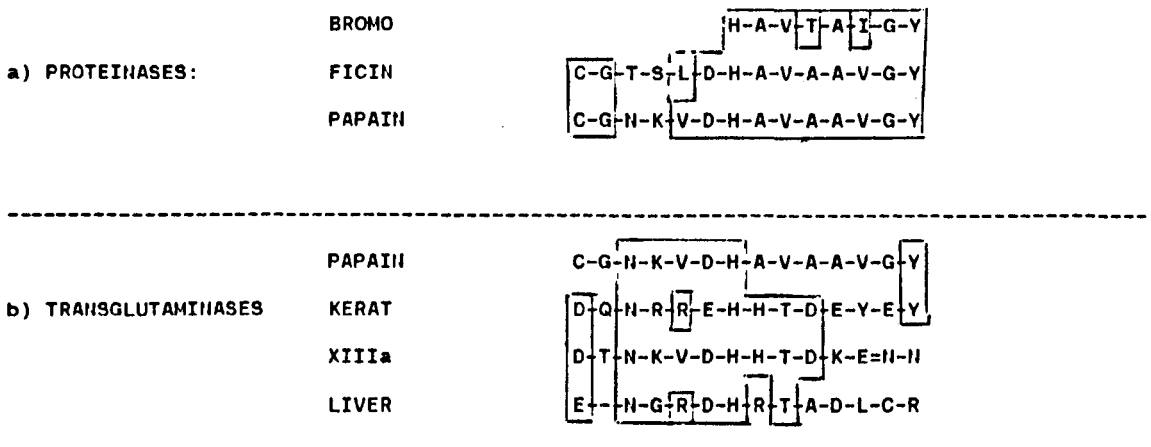
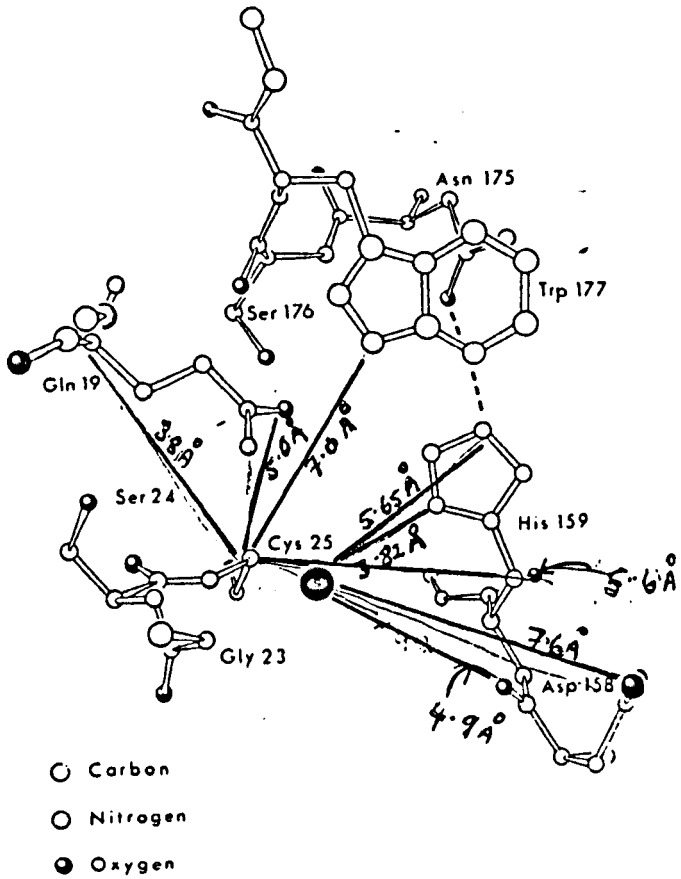


Fig 12.3. SEQUENCE ALIGNMENT FOR HOMOLOGY BETWEEN TRANSGLUTAMINASES AND THE HIS_{159} CONTAINING REGION OF THE ACTIVE SITE REGION OF PAPAIN Identical and related amino acid residues are boxed

The transglutaminases have greater similarity to papain than to other cysteine proteinases about this sequence. The consensus sequence may reduce to N-(A/G)-(I/V)-(E/D)-H . It is tempting to speculate that residues Asn_{16}, Asp_{19} and His_{20} in tgase may be homologous to residues Asn_{155}, Asp_{158} and His_{159} of papain (papaya) and project into the active site of transglutaminase. These residues could then be directly involved with the mechanism of catalysis of this family of enzymes. A diagram which shows the location of both the residues associated with the His and Cys motif of papain at its active site is shown in Fig 12.4.



12.4 THE ORIENTATION OF AMINO ACID SIDE CHAINS AT THE ACTIVE SITE OF PAPAIN

The diagram is taken from Drenth et al., (1971). It shows the proximity of several residues which are thought to be important in the chemistry of the reaction of papain with its substrates to Cys_{25}. This residue is known to be important in the catalytic mechanism of papain and like Cys_{275} of transglutaminase forms a thioester with synthetic ester substrates as part of its catalytic mechanism. A number of the residues which are highly conserved amongst the cysteine proteinases are also highly conserved in transglutaminases. A common catalytic mechanism suggests that these common residues may form a similar conformation at the active site of transglutaminase as is shown in papain. The situation may be analogous to the common catalytic triad His, Ser, Asp in the serine proteinases. It is worth noting that proteinases may be encouraged to catalyse the reverse reaction (ie. transpeptidation) under forcing conditions. This reaction is of the type catalysed by transglutaminases.

The interatomic distances were calculated with the assistance of Dr C Johnston from a computer graphical model of papain. Interestingly papain has been crystallized in some cases with the synthetic substrate CBZ-Gln-Gly ester which is a substrate for transglutaminase. In the model the CBZ- group does not appear to make any specific interactions with the active site of papain and lies predominantly along the external lip of the active site. This group probably increases the substrate effectiveness of the substrate by either blocking the positive charge of the amine terminus or increasing the hydrophobicity of the substrate or both. The same is probably true for the binding of this and related peptides at the active site of transglutaminase.

CHAPTER THIRTEEN: CONCLUSIONS AND FUTURE WORK

13.1. CONCLUSIONS

Early observations concerning the specificity of transglutaminase towards denatured protein substrates indicated that a wide variety of amino-acid sequences were associated with substrate sites (Toda and Folk, 1969). Hence transglutaminase was not considered to have a high specificity with respect to the primary amino acid sequence which surrounds reactive glutaminyl \residues within protein substrates. The local conformation which surrounded glutaminyl residues in proteins appeared to dictate their reactivity, and it was suggested that transglutaminase favoured unfolded regions within proteins as substrates. The selectivity of transglutaminase towards protein substrates appeared to largely depend on the presence of "fully exposed glutaminyl side-chains" and the possibility of movement in the region surrounding such sites, although even these observations were conspicuously understated. Later studies of the substrate properties of synthetic peptide substrates showed that quite large differences in Km were possible for single amino acid substitutions within synthetic peptides of between 8 and 15 residues in length. It has not been recognised until recently that such peptides may have secondary structure in aqueous solution, and hence these peptides were considered to all have "random coil" conformations. The Km values of the liver enzyme towards these peptides were different in many cases to those determined for factor XIIIa, and this suggested (contrary to the work on denatured proteins) that some determinants of specificity must reside within the primary sequence of protein substrates. With an increased knowledge of the likely solution structures of many of the substrates of transglutaminase, it has been possible to group substrate sites

according to their folded conformations and construct the set of rules based on allowed occupancies of amino-acids at positions C-, and N-terminal to labelling sites within type A, and type B substrates (see section 10.1.). It appears that there may be restrictions on the allowed distribution of charged residues within each structural class of substrate. The nature of the allowed distribution may have been previously obscured when both type A and type B substrates were considered together.

Since the sequence at the active site of papain resembles a similar sequence in transglutaminase it is possible to produce a theory which will explain the observed specificity differences between folded and unfolded substrates. The rules concerning transglutaminase specificity towards protein substrates may not be entirely complete at present, but have been found to be adequate for the purposes of:

1) successfully predicting the lack of reactivity of all non-labelled glutaminyl sites within the globular proteins: Phosphoglycerate kinase (yeast), phosphoglycerate mutase (yeast), the $His_{388} \rightarrow Gln_{388}$ mutant of yeast PGK, ß-Bp crystallin, ß-lactoglobulin and possibly a number of other proteins including : bovine catalase, rabbit muscle aldolase etc [though reactive sites have not been yet proven by direct sequencing of labelled sites (see section 11.5)].

2) providing a suitable explanation for the lack of reactivity of a number of exposed glutaminyl residues within denatured proteins.

3) explaining the observed labelling stoichiometries of a number of denatured proteins eg pepsin pH 8.0 and peptides, eg. the fifteen synthetic peptide substrates of Gorman and Folk, (1984) (see Appendix I).

4) Predicting a number of labelling sites within model systems which may be tested in order to further validate the rules themselves (eg. rabbit muscle aldolase is predicted to be modified at positions Gln_{155} and Gln_{345} if the rules are correct as they currently stand (see section 4.4.2.).

5) Providing an explanation for the differences in observed specificity between the guinea pig liver transglutaminase and factor XIIIa. **A slight difference in the distribution of charged residues at the active site of two transglutaminases could explain the differences in specificity towards the two enzymes if charged residues are important determinants of substrate effectiveness in proteins and peptides as has been suggested.**

It would clearly be of interest to test the validity of the rules by investigating the substrate properties of model peptides with positive and negatively charged residues at suitably favoured or disfavoured positions relative to a reactive glutaminyl residue. The secondary structures of such peptides could be tested by circular dichroism so that the appropriate set of rules, (ie. for <u>type A</u> or <u>type B</u> substrates) could be applied to the analysis of results.

13.2. THE FEASIBILITY OF ENGINEERING SUBSTRATE SITES FOR TRANSGLUTAMINASE INTO ANTIBODIES

As part of this thesis it has been shown possible to engineer a novel substrate site into proteins with retention of tertiary structure and functional properties eg. PGK $His_{388} \rightarrow Gln_{388}$ mutant (see chapter 6). There appears to be no *a priori* reason why other proteins may not be similarly mutated in order to introduce novel substrate sites for transglutaminase. The new site must be introduced into proteins at a position which is :

1) not crucial to the correct folding of the protein,

2) not near to substrate binding sites,

3) thought to influence the stability of the protein following translation.

This site must contain a highly exposed glutaminyl residue and it must obey the charge rules (see section 10.4.). Antibodies are highly folded proteins and the single example tested for its substrate characteristics, IgG B72.3 (see Table 4.2.) was unreactive towards transglutaminase, unless partly denatured. It may be difficult to introduce an appropriately sized loop, (associated with reactive sites in actin (see section 1.7.8.3.) into an antibody without affecting the functional properties of the protein. It was suggested that the molecule might be persuaded to refold by slow dialysis against a mixture of reduced and oxidized glutathione after labelling. However the regeneration of the original tertiary structure was

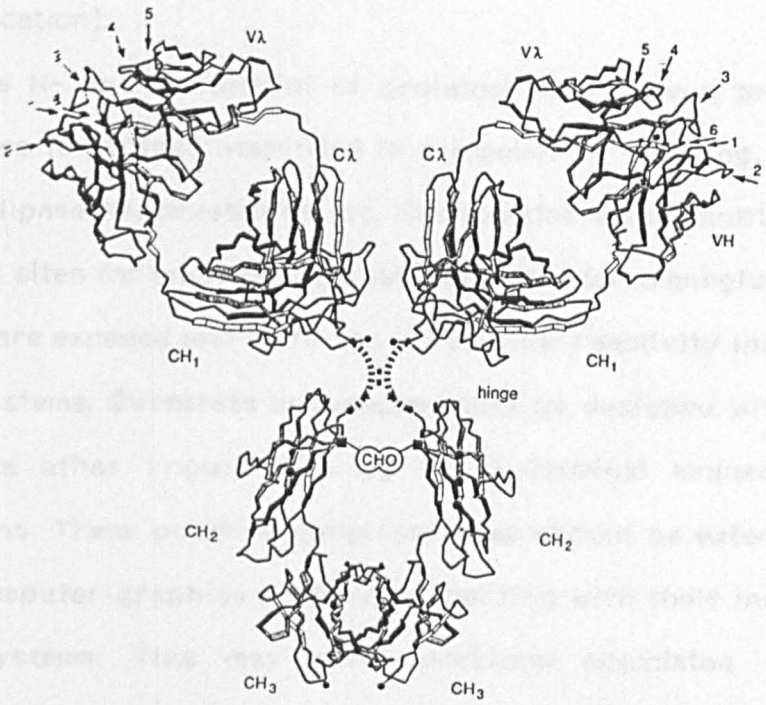


Fig 13.1. THE STRUCTURE OF AN ANTIBODY MOLECULE

The structure of a typical IgG molecule is shown schematically above. The antigen binding regions are located at the top of the diagram and the constant regions are located in the lower three quarters of the structure. The high ß-sheet content of the molecule is apparent and the individual strands are depicted as arrows. The diagram is from Winter and Milstein (1991).

The antibody IgG consists of four chains, two heavy and two light. These are constructed in a way that strings together domains of similar architecture. Each chain is paired with another chain by lateral packing of the domains and also by at least one disulphide bond. Each domain consists of two ß-sheets which pack together to form a sandwich, with exposed loops at the ends of the strands. Thus the C (constant) domains have three ß-strands in one sheet (strands C,F,G) and four strands (A, B, C, D) in the other. The framework is highly conserved in different antibodies and the individual domains resemble $ß_2$ microglobulin which has been shown to have substrate properties for transglutaminase (see Appendix I).

considered unlikely under these circumstances (Dr B. Smith personal communication).

The N- and C- termini of proteins have proven to have substrate properties for transglutaminase in a number of cases eg. ß-lactoglobulin, phospholipase A2, crystallins etc. Such region within antibodies represent potential sites for engineering substrate sites for transglutaminase as long as they are exposed and fulfil the criteria for reactivity indicated by other model systems. Substrate sequences could be designed with a motif which resembles other known sites eg the N-terminal sequences of the ß-crystallins. These putative substrate sites should be extensively modelled using computer graphics prior to proceeding with their incorporation into model systems. This may avoid problems associated with unforseen interactions of these sites with other regions of the molecule. Such interactions could affect exposure of the glutaminyl residue and cause a reduction in the substrate properties at this site.

13.3. PRODUCTION OF DESIRABLE IgG CONJUGATES

The specificity of transglutaminase for primary amines appears to be understood (see section 1.6.3.). Therefore it should be possible to attach covalently suitably long, unbranched, and uncharged amine groups to compounds so that transglutaminase may catalyse their attachment to reactive (possibly engineered) sites in proteins.

APPENDIX I APPLICATION OF THE CHARGE RULES TO PROTEIN SUBSTRATES OF TRANSGLUTAMINASE

I. INTRODUCTION

In chapter 10 the charge rules [developed to understand transglutaminase specificity] were applied to a number of proteins for which the modification sites were known, in order to test their predictive power. Since the rules were generally successful in the prediction of substrate sites within the model systems tested, their use has been extended in this appendix to the main work presented in this thesis. They are used here to:

1) predict the substrate sites within protein substrates of transglutaminase for which the modification sites are not presently known.

2) predict the local conformation associated with some of the known substrate sites of transglutaminase within portions of proteins which have not yet been structurally characterized.

In most cases the number of predicted substrate sites compare well with the maximum observed stoichiometry of modification of the proteins tested. Moreover where specific glutaminyl residues were strongly implicated on the basis of experimental studies using limited proteolysis these also were the residues predicted to have substrate properties by the rules.

II. THE PREDICTION OF LABELLING SITES WITHIN A NUMBER OF TRANSGLUTAMINASE SUBSTRATES

It should be noted that in no cases have proteins which have been characterized in terms of their reactivity or lack of reactivity towards transglutaminase and for which structural information is available been omitted from the discussion below.

III. HAEMAGGLUTININ

This was the only component of the intact virion to be labelled by transglutaminase, and was modified to a stoichiometry of approximately 1.7 mol putrescine/mol protein monomer (Iwanij, 1977). Freeze/thawing of the virions increased labelling to 3.0 mol putrescine/mol protein. Under normal conditions of maturation, the protein is cleaved into two polypeptides by the attack of plasmin (Lazarowitz, 1973). The use of other proteinases such as trypsin and chymotrypsin *in vitro* can generate similar peptides. The application of such proteinases to dansylcadaverine-labelled haemagglutinin quickly removed the label from the protein. Hence the labelling site appears to be associated with the loosely folded region which is accessible to both small enzymes such as trypsin and chymotrypsin and larger enzymes such as plasmin and transglutaminase (Iwanij, 1977).

The protein (Hong Kong H3N2 strain) has been crystallized and a X-ray structure is available at 3 Å resolution, (Wilson et al., 1981), see Fig 14.1. From this model it is possible to locate an appropriately "loosely folded regions" which may represent the reactive substrate site for transglutaminase. These are

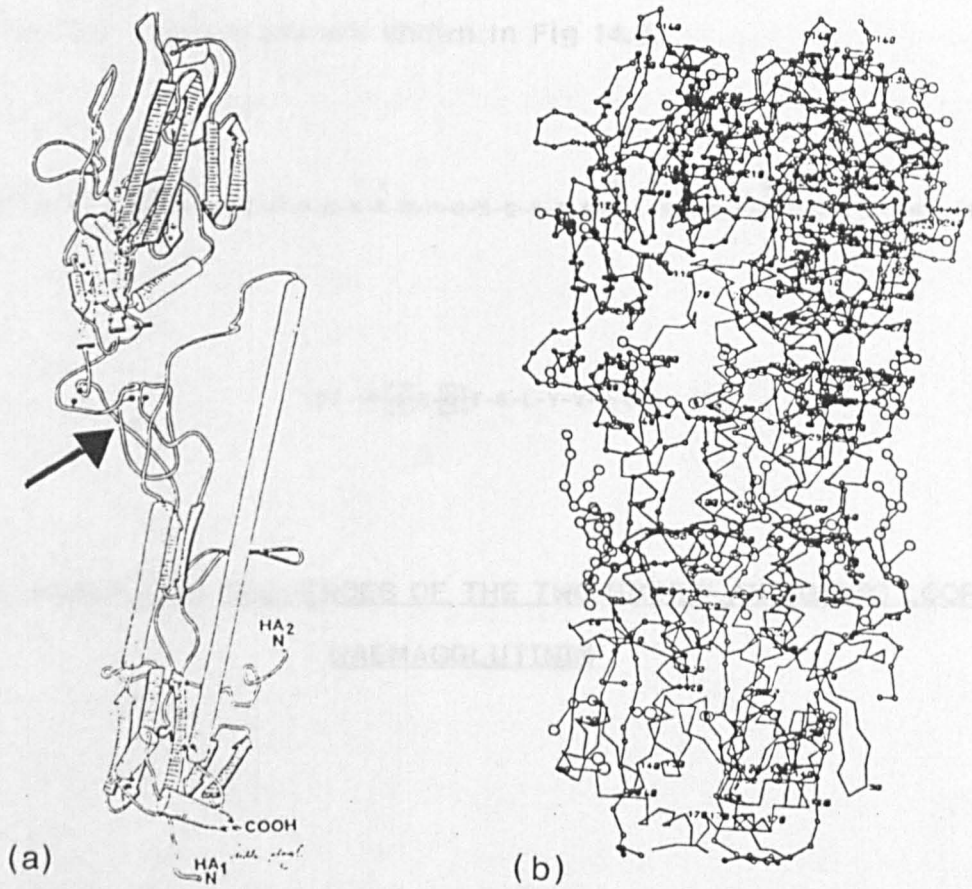


Fig 14.1. THE STRUCTURE OF HAEMAGGLUTTININ FROM INFLUENZA VIRUS

(a) Schematic drawing of the trace of the polypeptide in a subunit of haemagglutinin (Hong Kong strain H3N2) produced following bromelain digestion. The diagram is taken from Wilson et al., (1981) Regions of ß-sheet are flat twisted arrow-like bars and α-helix are tubes. The filled circles are disulphide bonds. The intact glycoprotein exists as a trimer shown in (b).
A number of loop structures (connecting strands) are surface exposed and are therefore particually susceptible to modification by 1) proteinases 2) transglutaminase. The loop structure which is implicated as a substrate of transglutaminase by virtue of its structure and the labelling studies of Iwanij (1980) is marked with an arrow.

located between residues 125-163 and 187-199. The conserved amino acid sequences within these areas are shown in Fig 14.2..

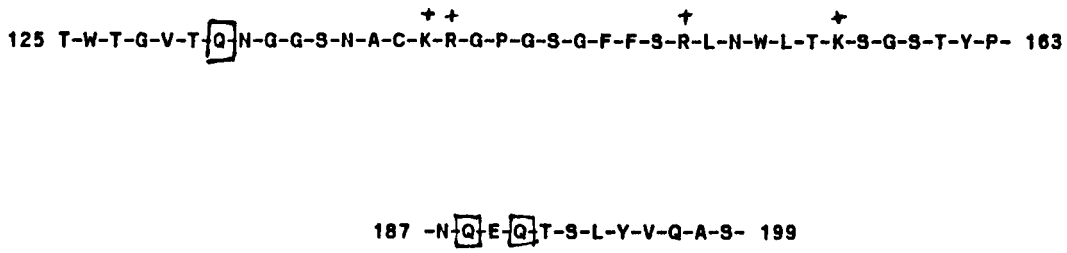


Fig 14.2. AMINO ACID SEQUENCES OF THE TWO HIGHLY ANTIGENIC LOOPS OF HAEMAGGLUTININ

The C-terminus of haemagglutinin HAI, [haemagglutinin primary structure is reviewed by Webster et al., (1982)] contains a highly conserved glutamine which could provide an appropriate substrate site in the intact HA molecule. The charge rules do not disallow any of the glutaminyl residues within the sequences shown in Fig 14.2. However steric limitations may restrict modification of residues at the ends of these surface loops, and by analogy with the topology of other known substrate sites Gln_{131} and Gln_{190} are predicted to have substrate properties for transglutaminase.

Iv HUMAN β-2-MICROGLOBULIN

Fesus et al. (1981) has reported incorporation of 1 mol/mol of methylamine

into this protein. Human β_2-microglobulin contains three glutaminyl residues, two of which (Gln_2, and Gln_8) are conserved in all species where the protein has been sequenced. The glutamine containing sequences are shown in Fig 14.3..

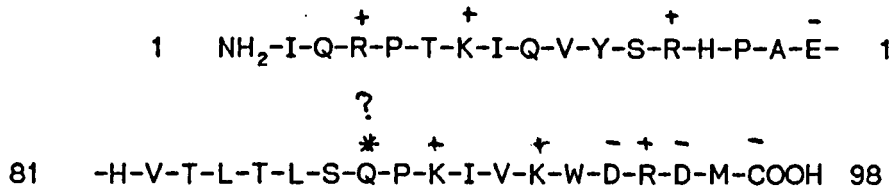
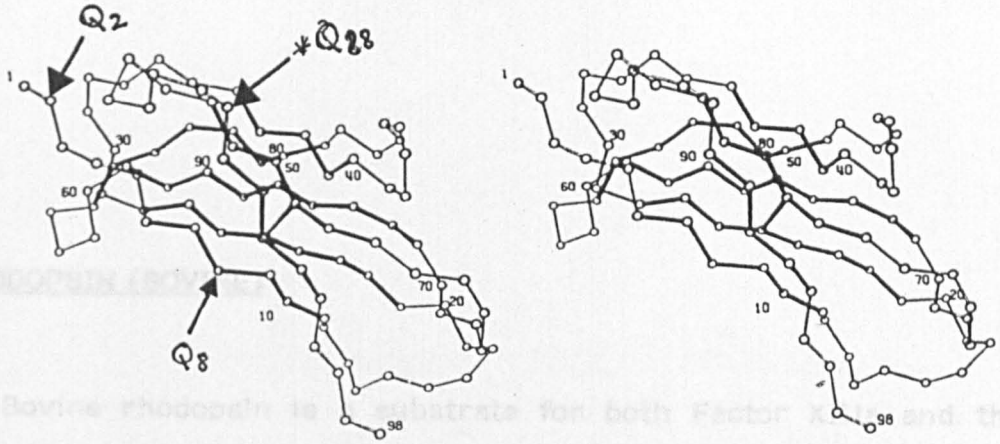
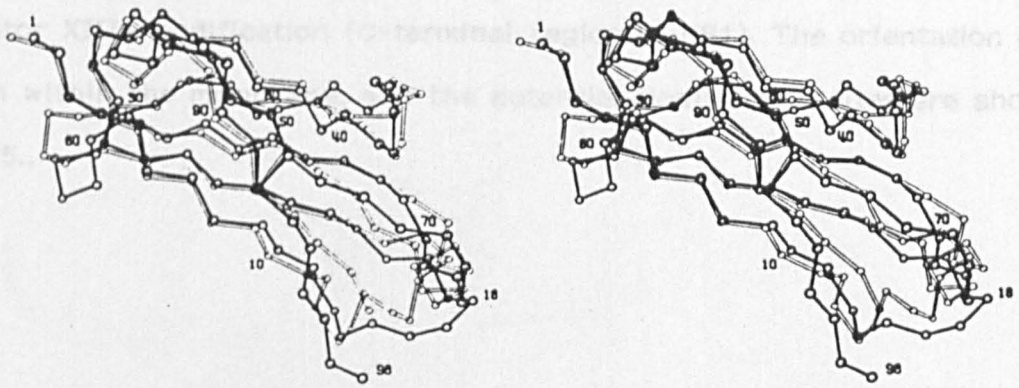


Fig 14.3 SEQUENCES SURROUNDING GLUTAMINES IN β_2-MICROGLOBULIN

The X-ray structure of β_2-microglobulin, (Becker and Reeke, 1985) is composed of a β-barrel configuration dominated by two antiparallel β-sheets, comprising three and four strands respectively, and is shown in Fig 14.4. None of the glutaminyl residues appear to be within these regions of β-sheet (residues 6 to 11, 21 to 28, 36 to 40, 49 to 56, 61 to 69, 77 to 82, 90 to 94). The glutaminyl residues are located on the N-terminal region, and within the C-terminus of the final loop structure of the protein. Map densities were weak in the areas 1 to 2, 41 to 45, 95 to 98. A disulphide bridge appears by analogy with the IgG structure to attach two of the β-sheet structures at positions Cys_{25} and Cys_{78} and is internal. Application of the charge rules suggests that only Gln_{88} should have substrate properties and only if a type B substrate. This is in accord with the observed stoichiometry of 1.0 mol methylamine/mol protein (Fesus et al., 1981).



(a)



(b)

Fig 14.4. THE STRUCTURE OF HUMAN β_2 MICROGLOBULIN

(a) Stereo drawing of the α-chain of α_2-microglobulin. Peptide bonds are represented by single lines peptide bonds in ß-structure and disulphide bonds are shown as darker lines. The view of the molecule is from down the crystallographic Y axis, with X horizontal and Z vertical. The structure is mainly composed of ß-sheet and has its tertiary fold in common with antibody IgG molecules. The position of glutaminyl residues are indicated by arrows. The diagram is taken from, Becker and Reeke (1985).

(b) A comparison of the tertiary structure of microglobulin with the constant region CH3 of human Fc (Deisenhofer et al., 1981).

v RHODOPSIN (BOVINE)

Bovine rhodopsin is a substrate for both Factor XIIIa and the liver enzyme (see section 1.8.2.). The modification sites have been determined for XIIIa using the succinylated substrate, (McDowell et al., 1986). The stoichiometry of modification observed using the liver enzyme was only 1.0 mol/mol, (Pober et al., 1978). In this case non-succinylated rhodopsin was used as substrate. The modification site(s) were shown to lie within the same external loop as observed for factor XIIIa modification (C-terminal region of FS1). The orientation of the protein within the membrane, and the potential modification sites are shown in Fig 14.5..

321	K-E-A-A-A-Q-Q-Q-E-S-A-T-T-Q-K-A-K-E	338
337	V-S-K-T-E-T-S-Q-V-A-P-A-COOH	348
61	V-T-V-Q-H-K-K-L-R-T-P-L-N-Y	74

Fig 14.5 POTENTIAL MODIFICATION SITES FOR TRANSGLUTAMINASE WITHIN RHODOPSIN

Application of the charge rules to bovine rhodopsin indicates that of the seven glutaminyl residues exposed at the aqueous face of the disc membranes, only three are compatible with the rules and these are present at positions Gln_{236}, Gln_{237}, and Gln_{238} (see Fig 14.6.). Since Pober et al., (1978) used both

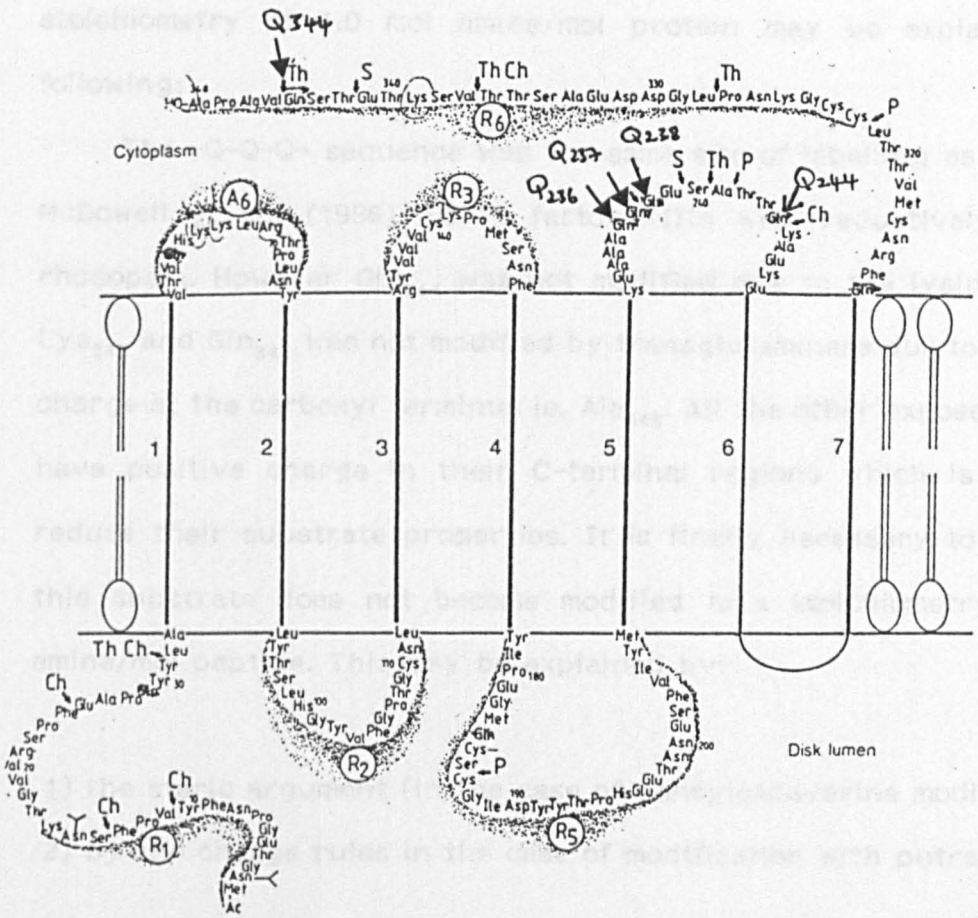


Fig 14.6. A MODEL OF THE MEMBRANE ORIENTATION OF BOVINE RHODOPSIN

Transglutaminase is thought to be able to gain access to the cytoplasmic side of the retinal discs. Hence the two loops (A6 and the region from Lys_{232} to Glu_{249}) extending from the lipid bilayer of the membrane together with the C-terminal portion (Gln_{312}-Ala_{340}) represent potential substrate sites for transglutaminase. The glutaminyl residues on the luminal side are not accessible for reaction. The residues known to be reactive towards factor XIIIa are indicated by arrows ↓. The Glutaminyl residue which is predicted to be the most reactive towards transglutaminase on the basis of the surrounding sequence is also indicated with a star. The model is from Ovchinnikov, (1987).

putrescine and dansylcadaverine to label the protein the observed stoichiometry of 1.0 mol amine/mol protein may be explained by the following:

The -Q-Q-Q- sequence was the same site of labelling as observed by McDowell et al., (1986). using factor XIIIa with reductively methylated rhodopsin. However Gln_{244} was not modified due to the lysine at position Lys_{245} and Gln_{344} was not modified by transglutaminase due to the negative charge at the carboxyl terminus ie. Ala_{348}. All the other exposed glutamines have positive charge in their C-terminal regions which is expected to reduce their substrate properties. It is finally necessary to explain why this substrate does not become modified to a stoichiometry of 3.0 mol amine/mol peptide. This may be explained by:

1) the steric argument (in the case of dansylcadaverine modification)
2) by the charge rules in the case of modification with putrescine.

McDowell et al., (1986) observed only 65 % modification of each of the glutamines in the -Q-Q-Q- sequence. The same degree of modification is considered likely in the case of the guinea pig liver modification. All three residues will have substrate properties, but after attachment of the first dansylcadaverine or putrescine molecule further modification of surrounding groups will not be possible. In the case of dansylcadaverine the large bulky group will not allow transglutaminase to approach closely enough to attach a second group. In the case of putrescine the modification of any glutamine will introduce positive charge which will inhibit further reaction of neighbouring glutamines over and above any effects caused by steric blocking.

vii MAJOR HISTOCOMPATIBILITY ANTIGENS

Pober and Strominger (1981) have demonstrated that 1 mol/mol of amine may be incorporated into the cytoplasmic C-terminus (Fig 14.7) of membrane bound HLA A2 and B7. The X-ray structure is determined for the extracelluar portion of this protein, (Bjorkman et al., 1987) but this does not include the reactive residues (see Fig 14.6.).

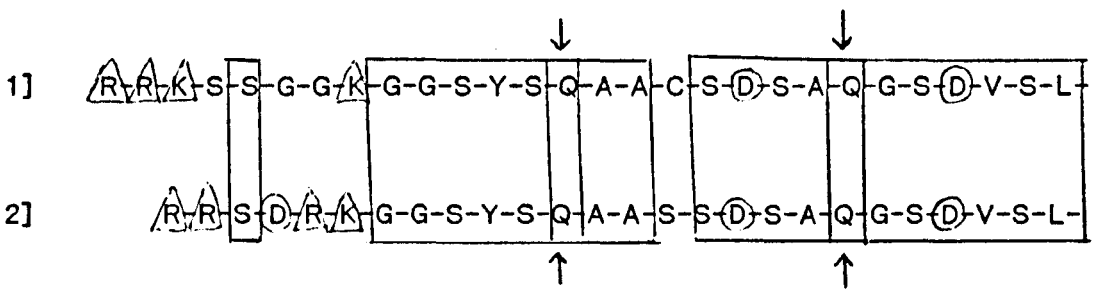


Fig 14.7. SEQUENCE OF C-TERMINUS OF [1] HLA A2 AND [2] HLA B7, ac-2. Positively charged residues are in triangles and negatively charged residues are in circles. Residues in eqivalent positions are boxed

The conformation of these sequences is not known, but may be quite flexible since these domains may be readily excised by limited proteolysis, (Guild and Strominger, 1984). The application of the rules indicates that both Gln_{322} and Gln_{330} of the C-termini of HLA-A2 and HLA-B7 are predicted to act as substrates for transglutaminase. However these sequences must be type A substrates for this to be allowed in both cases. If these sequences were considered type B, then both sites would be disallowed on the basis of the distribution of negative charge within the surrounding regions. That flexibility exists in this region is indicated by the

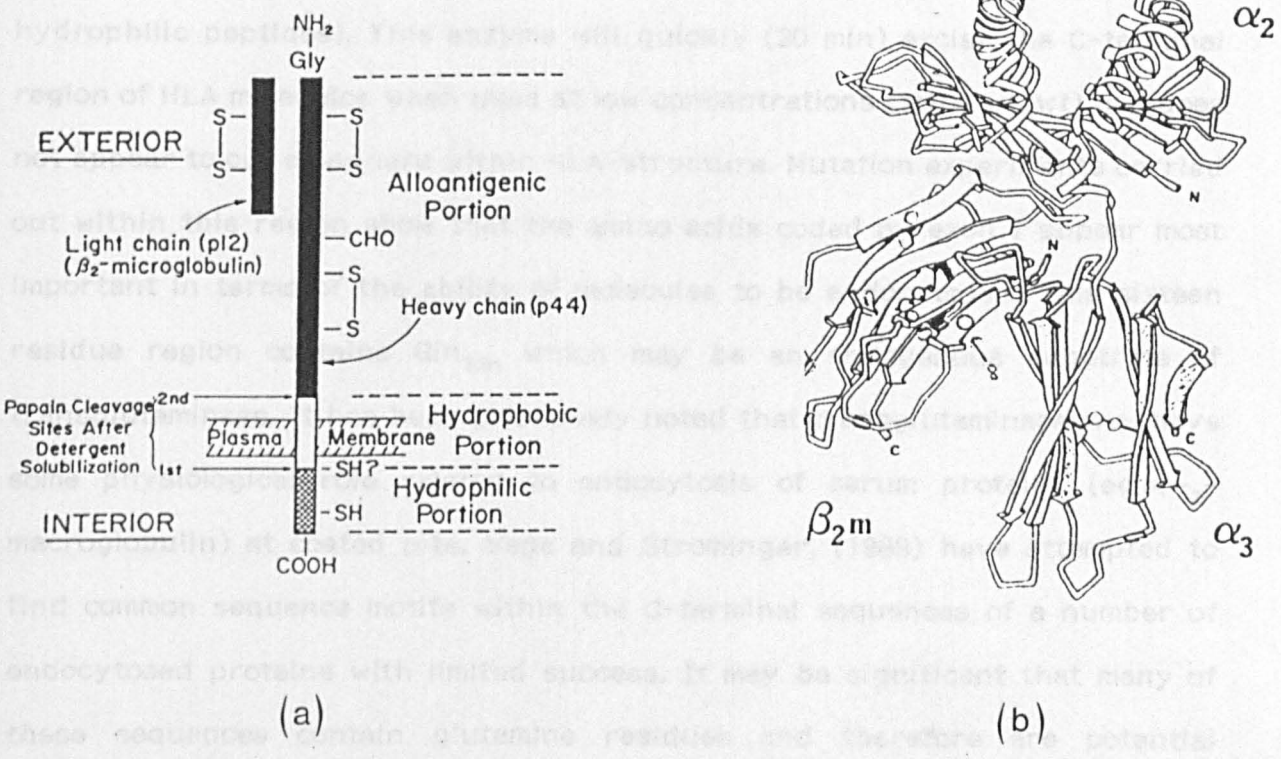


Fig 14.8. THE STRUCTURE OF THE HISTOCOMPATIBILITY ANTIGEN HLA-A2

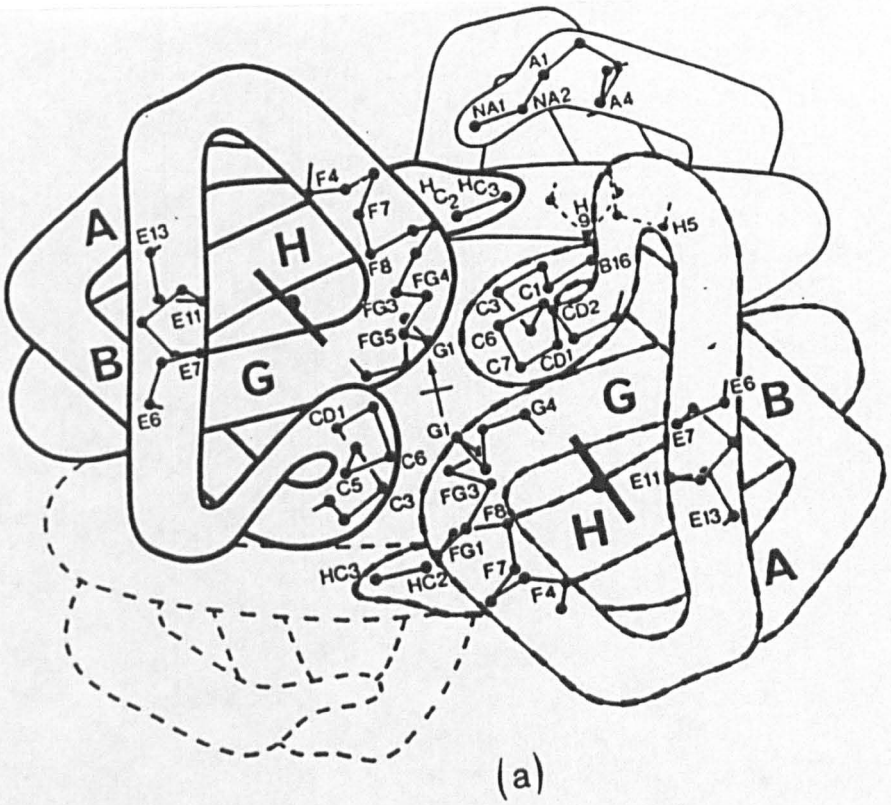
(a) Arrangement of the HLA- antigen in the membrane, with the hydrophillic peptide shown inside the cell membrane. The model is from Strominger et al., (1976).

(b)
Schematic drawing of the trace of the heavy and light chains of papain solubilized major histocompatibility antigen HLA A2/B7. Regions of ß-sheet are arrow-like bars and ∝-helix are tubes. The light chain ($ß_2$-microglobulin is depicted in association with the heavy chain. This is also a substrate for transglutaminase and is shown in Fig 14.4. The extracellular portion of the heavy chain of HLA-A2 resembles an antibody molecule in its domain structure. It extends into the membrane and terminates in a C-terminal cytoplasmic "tail" (see a). The structure of this part of the molecule is not known and this region has substrate properties for transglutaminase, while the extracellular portion does not. The diagram is taken from Bjorkman et al., (1987).

results of limited proteolysis of this region using *P.fragi* proteinase (which cleaves peptide bonds after aspartate and cysteine side chains in small hydrophilic peptides). This enzyme will quickly (30 min) excise the C-terminal region of HLA molecules when used at low concentrations (1:100 wt/wt) but does not appear to cut elsewhere within HLA-structure. Mutation experiments carried out within this region show that the amino acids coded by exon 7 appear most important in terms of the ability of molecules to be endocytosed. This sixteen residue region contains Gln_{330} which may be an endogenous substrate of transglutaminase. It has been previously noted that transglutaminase may have some physiological role related to endocytosis of serum proteins (eg α_2-macroglobulin) at coated pits. Vega and Strominger, (1989) have attempted to find common sequence motifs within the C-terminal sequences of a number of endocytosed proteins with limited success. It may be significant that many of these sequences contain glutamine residues and therefore are potential substrates of cytoplasmic transglutaminases.

vii HAEMOGLOBIN

From the X-ray structure of haemoglobin (see Fig 14.9) it is clear that the glutaminyl side chains are not sufficiently exposed to be highly effective substrates for transglutaminases. However it may be that the differences in labelling observed between horse and human haemoglobin (Brenner and Wold, 1978) reflect slight differences in substrate effectiveness and were not artifacts caused by the partial denaturation of samples. Application of the rules to the sequences surrounding the three glutaminyl residues located within the α and β chains of human haemoglobin indicates that Gln_{54} of the A chain should be a



(a)

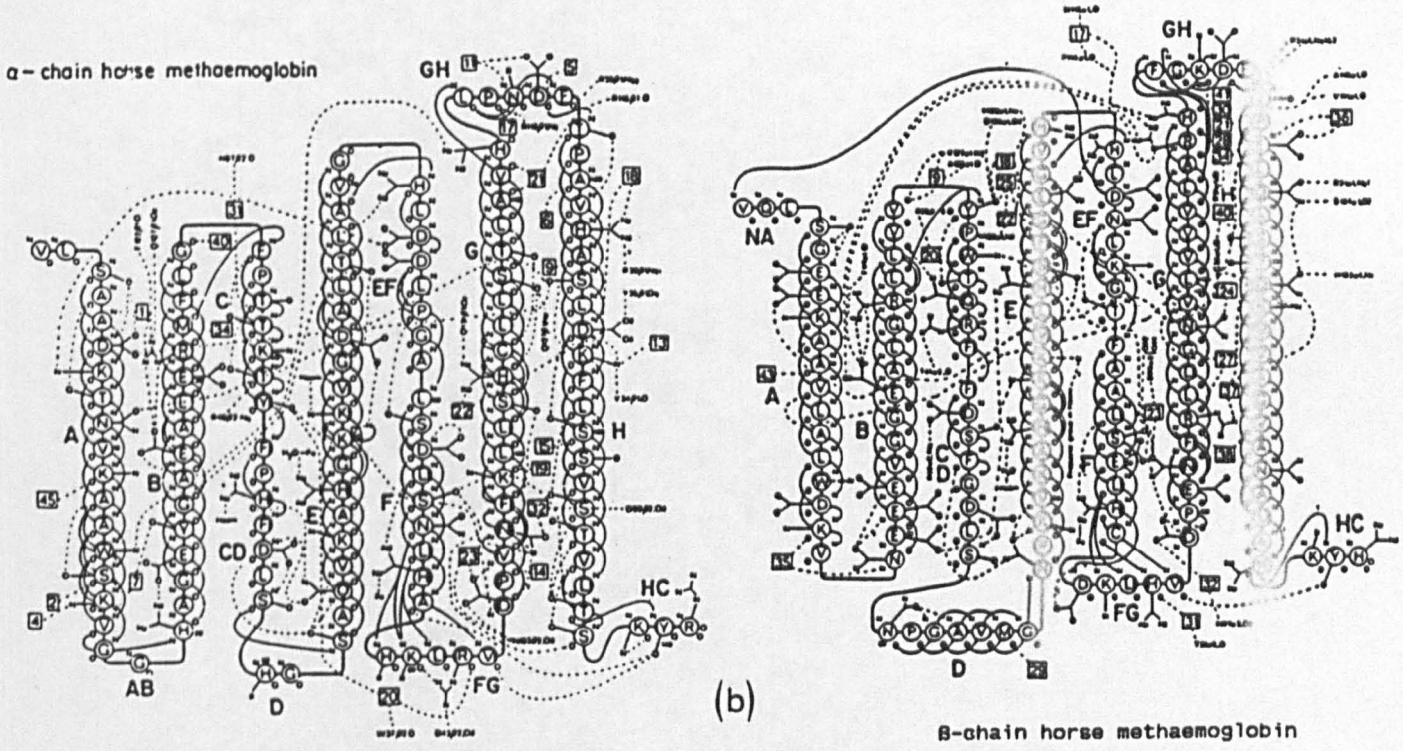


Fig 14.9. THE STRUCTURE OF HUMAN HAEMOGLOBIN

(a) The packing of α- and β- subunits into the haemoglobin tetramer.

(b) Although the diagrams are of horse haemoglobin, the sequence and tertiary fold are almost identical to that of the human protein. The network of hydrogen bonds stabilizing the structure of the α-chain and the β-chain of horse haemoglobin. Full lines represent covalent bonds, broken lines represent interactions involving side-chain atoms. Bonds in helices are shown as a half figure of eight. The diagrams are from Dickerson and Geis, 1983

Symbols: Open circles = external residues.

substrate (though a poor one) of transglutaminase when fully denatured. If this residue is suitably exposed, it may also be a substrate in the folded form. Glutaminyl residue Gln_{39} of the ß-chain would be disallowed on the basis of the positive charge at position +1 whether folded or denatured. Glutamine Gln_{127} is predicted to be a poor substrate (positive charge at position +5) in the denatured form and to not be a substrate in the folded form even if exposed (negative charge at position -6). Glutamine Gln_{131} is not predicted to be a substrate in any circumstances due to the presence of a positive charge at position +1. These predictions are in accord with the current state of knowledge of the 3-D structure, and the known sites of labelling of human haemoglobin.

PROTEINS AND PEPTIDES LOW IN SECONDARY STRUCTURE

viii. α-GLIADINS

The glutamine rich N-terminal region of α- A gliadin may be important in the promotion of coeliac disease in susceptible individuals. Conformational studies (c.d. spectroscopy) suggest that this region in A-gliadin is composed of type (I/III) ß-reverse turns (Tatham, et al., 1990). Similar structure was predicted using the method of Chou and Fasman (1974), (Tatham, et al., 1990) and this may explain the spiral structure observed using scanning tunnelling microscopy of the protein (Miles et al., 1991) . Although its substrate properties for transglutaminase have been noted (Porta et al., 1990), no figures relating to the stoichiometry of modification of the α-gliadins by transglutaminase were published. The primary amino acid sequence of the N-terminus of A-gliadin is shown in Fig 14.10.

$_{3}$ VPVPQLQPQNPSQQQPQ(E)QVPLVQQQQFLGQQQPFPPQQPYPQPQPFPSQQP$_{55}$

Fig 14.10 THE N-TERMINAL SEQUENCE OF A-GLIADIN (WHEAT)

The application of the rules to the N-terminal sequence of A-gliadin (Kasarda, 1980) suggests that many of these glutaminyl residues are predicted to have substrate properties both as either type A or type B substrates. Thus the substrate properties of this protein are in accord with those predicted by the rules.

ix. SYNTHETIC PEPTIDES

Since the synthetic peptides tested by Gorman and Folk, (1984) and Folk, (1983) were not included in the formulation of the original rules it is of interest to see how the changes in Km observed for a variety of amino acid substitutions within two synthetic model peptide substrates correlate with the rules, ie.

1) does the Km for a particular substrate rise when a substitution predicted to be unfavourable by the rules is tested?

2) Conversely; does the substitution of an amino acid side chain which is predicted to be unfavourable by the rules, for one which is not predicted to reduce substrate properties cause the expected decrease in Km.

Since all of the peptides tested were relatively short and did not have the type of charge distribution considered favourable for the stabilization of α-helical structure, it seems likely that most of these peptides were either lacking in folded secondary structure or were in rapid equilibrium between folded forms. Hence the rules governing flexible (type A) substrates would be expected to explain the observed pattern of reduction or increase in Km following substitution.

In most of the peptides tested as substrates, positive charge was located at position +2 which is predicted to be allowed but unfavoured. Thus the relatively high Km values obtained for the action of transglutaminase on these peptides (in the 1.0 to 5.0 mM range) can be explained. In no cases were residues in the C-terminal region to the glutamines substituted for positively charged residues, nor was a negative charge tested at position +4 relative to the reactive glutamine hence according to the rules it is not surprising that single substitutions of Gly, Ala or Ser within these peptides had little obvious effects on transglutaminase specificity. Indeed cumulative substitutions of up to three Ser residues within regions allowed by the rules had little effect on the specificity of the enzyme towards such substrates. These observations are in accord with the specificity rules for type A substrates.

Gorman and Folk, (1984) noted that N-terminally succinylated derivatives of peptides containing a positively charged lysine residue at position +2 (relative to the glutaminyl residue) are not better substrates than the non-succinylated analogues and that this may be due to the unfavourable positive charge associated with the lysine residue. However when the N-terminal residue

of a peptide lacking in positive charge at this position (Lys to Ala substitution) was N-succinylated, its substrate properties were seen to improve with this modification. Moreover bis-succinylation of both the N-α-terminus and the ϵ-NH_2 of the lysyl- amine group increased substrate properties despite the +VE→ -VE charge reversal and the increase in steric hinderance about the reactive glutaminyl residue. These findings are also in accordance with the rules.

The overall results of the work done on model peptides are not inconsistent with the determinants of transglutaminase specificity which have been suggested in chapter 10. It appears that the presence of hydrophobic groups also favours reactivity of such model substrates, particularly in the region C-terminal to the glutamine and this is in accordance with the likely hydrophobic nature of the active site of the enzyme.

xi PORCINE PEPSIN

OXIDIZED PEPSIN

Porcine pepsin has been sequenced from the protein (Tang et al.,(1973), and an X-ray structure (see Fig 8.5.) is available for the protein (Andreeva et al., 1977). It is evident from the c.d. spectra and the modification data that most of the secondary structure of pepsin is retained at pH 8.0 in the absence of reducing agents (see section 8.3.2.1.). From modification experiments 2.0 mol putrescine/mol pepsin was observed to be the maximum stoichiometry of labelling of the oxidized protein.

The X-ray structure of pepsin shows that the protein is extensively folded into a bilobed globular structure, largely composed of β-sheet, β-

turn and loop structures. Glutaminyl residues Gln_{211}, Gln_{277} and Gln_{54}, are located within +1,+4, and +5, residues respectively from disulphide bonds. The conformational restriction due to these bonds in other proteins (Toda and Folk, 1959) appears to reduce substrate properties of adjacent glutaminyl residues. Glutaminyl residue Gln_{266} is located midway between the disulphide bond formed between Cys_{249} and Cys_{282} and therefore may also be more conformationally restricted than glutaminyl residues which are not located within these structures. In the absence of DTT approximately three fewer sites were available compared to the reduced protein indicating a higher degree of folding in the oxidized protein than may simply be explained by steric blockage of glutaminyl residues in the immediate vicinity of the disulphide bridges. Although most of the glutaminyl residues within pepsin are located within regions where steric restrictions would be expected to preclude their reaction with transglutaminase, the glutaminyl residues in the N- and C- termini ie Gln_{25} and Gln_{308} may have sufficient local flexibility at pH 8.0 to have substrate properties especially since the surrounding residues do not contravene the charge rules described in chapter 10.

REDUCED PEPSIN

Following the incubation of porcine pepsin with DTT the maximum stoichiometry of modification the protein by transglutaminase was raised to 6.0 mol putrescine/mol protein. The increase in the substrate properties of the protein are explicable on the basis of the ability of DTT to cleave one or more disulphide bonds within the molecule, followed by a loosening of the proteins structure and the exposure of reactive glutaminyl residues

The sequence of porcine pepsin is shown in appendix III (ix). Inspection of the sequence indicates that the protein is extremely rich in negative charge, and contains few positively charged side chains. Thus the lack of reactivity of glutaminyl side chains which have reduced substrate properties in the presence of DTT may not be explained simply on the basis of the inhibitory action of local positively charged residues. These residues must therefore either be sterically hindered from reaction, or the presence of specific residues other than Lys and Arg eg. certain negatively charged residues may be inhibitory to reaction in particular positions relative to glutaminyl residues (see section 10.4.).

Application of the charge rules, indicates that Gln_{55} would not be expected to have substrate properties if considered either type A or type B substrate. Independently of the restrictions imposed on substrates by the charge rules, other amino acid side chains have been noted as potentially unfavourable to reactivity of glutaminyl residues eg. cysteine, tryptophan phenylalanine etc. (see section 10.2.2.).

There are a total of thirteen glutaminyl residues within porcine pepsin and application of the specificity rules described in chapter 10 for either type A or type B substrates at each site predicts that seven and ten sites respectively should be reactive in each case. Most of the sites would be expected to be type A under reducing conditions and hence the predicted number of sites compares reasonably with the experimentally determined maximum stoichiometry of 6.0 mol dansylcadaverine/mol pepsin. The single modification site determined for transglutaminase within reduced pepsin ie. Gln_{308} (see section 8.3.2.2.) obeys the rules if considered a type A (flexible) substrate which is reasonable given the partially denatured state of the protein during its modification by transglutaminase.

xii OTHER PROTEINS TESTED AS PART OF THIS THESIS

A number of other substrates of transglutaminase for which maximum labelling stoichiometries are known, have high resolution crystal structures available and their substrate properties have been considered in terms of the exposure of their glutaminyl residues elsewhere in this thesis. In the case of aldolase (see section 4.4.2.) application of the rules predicts the same number of potentially reactive sites as are predicted purely on the basis of exposure of side-chains ie 2/monomer as is observed. In the case of catalase (see section 4.4.3.) five glutaminyl residues appear to be sufficiently exposed for transglutaminase to modify them. The charge rules predict that two of the exposed residues will not be efficient substrates. This reduces the number of potentially reactive sites to three, which is observed to be the maximum labelling stoichiometry.

xiii SUMMARY

The partially characterized substrates of transglutaminase considered in sections iii. to xii. show similar patterns of reactivity to that seen in the better characterized substrates which are discussed in section 1.7.. In some cases substrate sites must be strictly defined as either type A or type B in order to understand why charged residues neighbouring a glutaminyl residue do not cause the glutaminyl site to be unreactive towards transglutaminase.

APPENDIX II

I ASSAYS

1) 3-PHOSPHOGLYCERATE KINASE (PGK)

PRINCIPLE

(1) glycerate-3-P + ATP $\rightleftharpoons$ glycerate-1,3-P_2 + ADP

(2) glycerate-1,3-P_2 + NADH + H^+ $\rightleftharpoons$ GAP + NAD^+ + P

The change in optical density at 340 nm per unit time is a measure of the PGK activity.

Stock solutions

I. Triethanolamine buffer, 0.1 M, pH 7.6 (1.86g triethanolamine hydrochloride/100 ml, adjusted to pH 7.6 with 1N NaOH)

II. ATP solution, 32 mM (20 mg ATP,Na salt/ml).

III. Glycerate-3-P soln, 93 mM (50 mg glycerate-3-P, CHA-salt/ml).

IV. NADH soln, approximately 12 mM (10 mg NADH, Na-salt/ml).

V. EDTA soln, 27 mM [10 mg. EDTA/ml triethanolamine buffer (I.)]

VI. Magnesium sulphate solution, 0.1 M (2.47g $MgSO_4$. 7 H_2O/100 ml).

VII. Glyceraldehyde-3-phosphate dehydrogenase, from rabbit muscle, GAP-DH, crystalline suspension in ammonium sulphate solution (10 mg/ml), approx. 80 U/ml

Sample. [10 mg/ml PGK (450 U/mg) diluted 1:5000 with ice cold triethanolamine buffer (I.) immediately before the measurement.

Assay mixtures

To 2.5 ml of solution (I.) add: 0.1ml solution (II.)
0.2 ml solution (III.)
0.05 ml solution (IV.)
0.1 ml solution (V.)
0.06 ml solution (VII.)
0.01 ml solution (VII.)

Mix the ingredients by inversion of the tube, then add 0.02 ml of an appropriately diluted sample containing the enzyme. Read the optical density after 1, 2, 3, and 4 min and calculate the average change in OD_{340}/min.

Calculation

$$\text{Volume activity (U/ml)} = \frac{3.04}{6.22 \cdot 1.0 \cdot 0.02} \times \text{Change in OD/min}$$

$$\text{Specific activity (U/mg)} = \frac{\text{Volume activity}}{\text{Protein concentration}}$$

2) CATALASE

Principle

$$2\ H_2O_2 \xrightarrow{\text{catalase}} 2\ H_2O + O_2$$

The change in optical density at 240 nm per unit time is a measure of the activity of catalase activity.

Solutions

i) Phosphate buffer 50 mM, pH 7.0

ii) Buffer/H_2O_2 add to 50 ml of buffer i) 0.06 ml of 30% H_2O_2 and mix. Adjust the optical density of the solution to 0.5 at 240 nm (d = 1 cm).

Assay system

To 3 ml of buffer ii, add 20 ul of an appropriately diluted sample of catalase. Mix and measure how long it takes for the $O.D._{240}$ to drop from 0.45 to 0.40. The amount of enzyme should then be adjusted for tis to take 20 s.

The volume activity may then be calculated and the specific activity is the volume activity divided by the protein concentration.

$$\text{Volume activity} = \frac{17 \times 13.1}{\text{time (s)} \times 0.02}$$

PAGE MISSING IN ORIGINAL

APPENDIX III

GOOD SUBSTRATES FOR TRANSGLUTAMINASE

I. RABBIT ACTIN (MUSCLE)

Ac-D-E-T-E-D-T-A-L-V-C-D-G-S-G-L-V-K-A-G-F-A-G-D-D-A-P-R-A-V-F-P-S-I-V-G-R-P-R-H-Q-G-V-M
-V-G-M-G-Q-K-D-S-Y-V-G-D-E-A-Q-S-K-R-G-I-L-T-L-K-Y-P-I-E-X-W-G-I-I-T-N-D-D-M-E-K-I-W-H-H
-T-F-Y-N-E-L-R-V-A-P-E-E-H-P-T-L-L-T-E-A-P-L-N-P-K-A-N-R-E-K-M-T-Q-I-M-F-E-T-F-N-V-P-A-M
-Y-V-A-I-Q-A-V-L-S-L-Y-A-S-G-R-T-T-G-I-V-L-D-S-G-D-G-V-T-H-N-V-P-I-Y-E-G-Y-A-L-P-H-A-I-M
-R-L-D-L-A-G-R-D-L-T-D-Y-L-M-K-I-L-T-E-R-G-Y-S-F-V-T-T-A-E-R-E-I-V-R-D-I-K-E-K-L-C-Y-V-A
-L-D-F-E-N-E-M-A-T-A-A-S-S-S-L-E-K-S-Y-E-L-P-D-G-Q-V-I-T-I-G-N-E-R-F-R-C-P-E-T-L-F-Q-P-S
-F-I-G-M-E-S-A-G-I-H-E-T-T-Y-N-S-I-M-K-C-D-I-D-I-R-K-D-L-Y-A-N-N-V-M-S-G-G-T-T-M-Y-P-G-T
-A-D-R-M-Q-K-E-I-T-A-L-A-P-S-T-M-K-I-K-I-I-A-P-P-E-R-K-Y-S-V-W-I-G-G-S-I-L-A-S-L-S-S-T-F
-Q-Q-M-W-I-T-K-Q-E-Y-D-E-A-G-P-S-I-V-H-R-K-C-F

Sequence from : Woodrum, D.T. Rich, S.A., and Pollard, T. D. J. Cell Biol. 67 , 231 -237 (1975).

ii. RABBIT ALDOLASE (MUSCLE)

P-H-S-H-P-A-L-T-P-E-Q-K-K-E-L-S-D-I-A-H-R-I-V-A-P-G-K-G-I-L-A-A-D-Q-S-T-G-S-I-A-K-R-L-Q-
S-I-G-T-E-N-T-E-E-N-R-R-F-Y-R-Q-L-L-L-T-A-D-D-R-V-N-P-C-I-G-G--V-I-L-F-H-E-T-L-Y-Q-K-A-D
-D-G-R-P-F-P-Q-V-I-K-S-K-G-G-V-V-G-I-K-V-D-K-G-V-V-P-L-A-G-T-D-G-E-T-T-T-Q-G-L-D-G-L-S-E
-R-C-A-Q-Y-K-K-D-G-A-D-F-A-K-W-R-C-V-L-K-I-G-Q-H-T-P-S-A-L-A-I-M-E-N-A-N-V-L-A-R-Y-A-S-I
-C-Q-Q-N-G-P-I-E-V-P-E-I-L-P-D-G-D-H-D-L-K-R-C-Q-Y-V-T-Q-K-V-L-A-A-V-Y-K-A-L-S-N-H-H-I-Y
-L-Q-G-T-L-L-K-P-N-H-V-T-P-G-H-A-C-T-Q-K-Y-S-H-Q-Q-I-A-M-A-T-V-T-A-L-R-R-T-V-P-P-A-V-T-G
-V-T-F-L-S-G-S-E-E-E-E-G-A-S-I-N-L-N-A-I-N-K-C-P-L-L-W-P-K-A-L-T-F-S-Y-G-R-A-L-Q-A-S-A-L
-K-A-W-G-G-K-K-E-N-L-K-A-A-Q-E-E-Y-V-K-R-A-L-A-N-S-L-A-C-Q-G-K-Y-T-P-S-G-Q-A-G-A-A-A-S-E
-S-L-F-I-S-N-H-A-Y

Sequence from : Tolan, D.R., Amsden, A.B., Putney, S.D., Urdea, M.S., and Penhoet, E.E. (1984) J. Biol. Chem. 259 , 1127 - 1131.

iii. HUMAN LIPOCORTIN 1

H-A-M-V-S-E-F-L-K-Q-A-W-F-I-E-N-E-E-Q-E-Y-V-Q-T-V-K-S-S-K-G-G-P-G-S-A-V-S-P-Y-P-T-F-N-P-
S-S-D-V-A-A-L-H-K-A-I-M-V-K-G-V-D-E-A-T-I-I-D-I-L-T-K-R-N-N-A--Q-R-Q-Q-I-K-A-A-Y-L-Q-E-T
-G-K-P-L-D-E-T-L-K-K-A-L-T-G-H-L-E-E-V-V-L-A-L-L-K-T-P-A-Q-F-D-A-D-E-L-R-A-A-M-K-G-L-G-T
-D-E-D-T-L-I-E-I-L-A-S-R-T-N-K-E-I-R-D-I-N-R-V-Y-R-E-E-L-K-R-D-L-A-K-D-I-T-S-D-T-S-G-D-F
-R-N-A-L-L-S-L-A-K-G-D-R-S-E-D-F-G-V-N-E-D-L-A-D-S-D-A-R-A-L-Y-E-A-G-E-R-R-K-G-T-D-V-N-V
-F-N-T-I-L-T-T-R-S-Y-P-Q-L-R-R-V-F-Q-K-Y-T-K-Y-S-K-H-D-M-N-K-V-L-D-L-E-L-K-G-D-I-E-K-C-T
-T-A-I-V-K-C-A-T-S-K-P-A-F-F-A-E-K-L-H-Q-A-M-K-V-G-V-T-R-H-K-A-L-I-R-I-M-
V-S-R-S-E-I-D-M-N-D-I-K-A-F-Y-Q-K-M-Y-G-I-S-L-C-Q-A-I-L-D-E-K-G-D-Y-E-K-I-L-V-A-L-C-G-G-N

Sequence from : Wallner, B.P., Mattaliano, R.J., Hession, C., Cate, R.L., Tizard, R., Sinclair, L.K., Foeller, C., Chow, E.P., Browning, J.L., Ramachandran, K.L., and Pepinsky, R.B. (1986) Nature, 320 , 77 - 80.

Iv. PORCINE PANCREATIC PHOSPHOLIPASE A2

A-L-W-Q-F-R-S-M-I-K-C-A-I-P-G-S-H-P-L-M-D-F-N-N-Y-G-C-Y-C-G-L-G-G-S-G-T-P-V-D-E-L-D-R-C-C-E-T-H-D-N-C-Y-R-D-A-K-N-L-D-S-C-K-F-L-V-D-N-P-Y-T-E-S-Y-S-Y-S-C-S-N-T-E-I-T-C-N-S-K-N-N-A-C-E-A-F-I-C-N-C-D-R-N-A-A-I-C-F-S-K-A-P-Y-N-K-E-H-K-N-L-D-T-K-K-Y-C

Sequence from : Waite, M. (1987) *The phospholipases*: Handbook of lipid research, Vol 5. Plenum Publishing corp., New York.

v. PHOSPHOGLYCERATE KINASE (YEAST) HIS_{388} GLN_{388} MUTANT

S-L-S-S-K-L-S-V-Q-D-L-D-L-K-D-K-R-V-F-I-R-V-D-F-N-V-P-L-D-G-K-K-I-T-S-N-Q-R-I-V-A-A-L-P-
T-I-K-Y-V-L-E-H-H-P-R-Y-V-V-L-A-S-H-L-G-R-P-N-G-E-R-N-E-K-Y-S--L-A-P-V-A-K-E-L-Q-S-L-L-G
-K-D-V-T-F-L-N-D-C-V-G-P-E-V-E-A-A-V-K-A-S-A-P-G-S-V-I-L-L-E-N-L-R-Y-H-I-E-E-E-G-S-R-K-V
-D-G-Q-K-V-K-A-S-K-E-D-V-Q-K-F-R-H-E-L-S-S-L-A-D-V-Y-I-N-D-A-F-G-T-A-H-R-A-H-S-S-M-V-G-F
-D-L-P-Q-R-A-A-G-F-L-L-E-K-E-L-K-Y-F-G-K-A-L-E-N-P-T-R-P-F-L-A-I-L-G-G-A-K-V-A-D-K-I-Q-L
-I-D-N-L-L-D-K-V-D-S-I-I-I-G-G-G-M-A-F-T-F-K-K-V-L-E-N-T-E-I-G-D-S-I-F-D-K-A-G-A-E-I-V-P
-K-L-N-E-K-A-K-A-K-G-V-E-V-V-L-P-V-D-F-I-I-A-D-A-F-D-A-S-A-N-T-K-T-V-T-D-K-E-G-I-P-A-G-W
-Q-G-L-D-N-G-P-E-S-R-K-L-F-A-A-T-V-A-K-A-K-T-I-V-W-N-G-P-P-G-V-F-E-F-E-K-F-A-A-G-T-K-A-L
-L-D-E-V-V-K-S-S-A-A-G-N-T-V-I-I-G-G-D-T-A-T-V-A-K-K-Y-G-V-T-D-K-I-S-H-V-S-T-G-G-G-A-S-L
-E-L-L-E-G-K-E-L-P-G-V-A-F-L-S-E-K-K

Sequence from : Hitzeman, R.A., Hagie, F.E., Hayflick, J.S., Chen C.Y., Seeburg, P.H., and Derynk, R. (1982) Nucleic acid Res. 10 , 7791 - 7808.

vi. BOVINE β-CASEIN

R-E-L-E-E-L-N-V-P-G-E-I-V-E-S-L-S-S-S-E-E-S-I-T-R-I-N-K-K-I-E-K-F-Q-S-E-E-Q-Q-Q-T-E-D-E-
L-Q-D-K-I-H-P-F-A-Q-T-Q-S-L-V-Y-P-F-P-G-P-I-P-N-S-L-P-Q-N-I-P--P-L-T-Q-T-P-V-V-V-P-P-F-L
-Q-P-E-V-M-G-V-S-K-V-K-E-A-M-A-P-K-H-K-E-M-P-F-P-K-Y-P-V-Q-P-F-T-E-S-Q-S-L-T-L-T-D-V-E-N
-L-H-L-P-P-L-L-Q-S-W-M-H-Q-P-H-Q-P-L-P-P-T-V-M-F-P-P-Q-S-V-L-S-L-S-Q-S-K-V-L-P-V-P-Q-K
-A-V-P-Y-P-Q-R-D-M-P-I-Q-A-F-L-L-Y-Q-Q-P-V-L-G-P-V-R-G-P-F-P-I-I-V

Sequence from : Holt, C., and Sawyer, L. (1988) Protein Engineering 2 , (4) 251 -259.

vii. BOVINE RHODOPSIN

M-N-G-T-E-G-P-N-F-Y-V-P-F-S-N-K-T-G-V-V-R-S-P-F-E-A-P-Q-Y-Y-L-A-E-P-W-Q-F-S-M-L-A-A-Y-M-
F-L-L-I-M-L-G-F-P-I-N-F-L-T-L-Y-V-T-V-Q-H-K-K-L-R-T-P-L-N-Y-I-L-L-N-L-A-V-A-D-L-F-M-V-F-
G-G-F-T-T-T-L-Y-T-S-L-H-G-Y-F-V-F-G-P-T-G-C-N-L-E-G-F-F-A-T-L-G-G-E-I-A-L-W-S-L-V-V-L-A-
I-E-R-Y-V-V-V-C-K-P-M-S-N-F-R-F-G-E-N-H-A-I-M-G-V-A-F-T-W-V-M-A-L-A-C-A-A-P-P-L-V-G-W-S-
R-Y-I-P-E-G-M-Q-C-S-C-G-I-D-Y-Y-T-P-H-E-E-T-N-N-E-S-F-V-I-Y-M-F-V-V-H-F-I-I-P-L-I-V-I-F-
F-C-Y-G-Q-L-V-F-T-V-K-E-A-A-A-Q-Q-Q-E-S-A-T-T-Q-K-A-E-K-E-V-T-R-M-V-I-I-M-V-I-A-F-L-I-C-
W-L-P-Y-A-G-V-A-F-Y-I-F-T-H-Q-G-S-D-F-G-P-I-F-M-T-I-P-A-F-F-A-K-T-S-A-V-Y-N-P-V-I-Y-I-M-
M-N-K-Q-F-R-N-C-M-V-T-T-L-C-C-G-K-N-P-L-G-D-D-E-A-S-T-T-V-S-K-T-E-T-S-Q-V-A-P-A

Sequence from : Nathans, J. and Hogness, D.S. (1983) Cell 34 , 807 - 814.

viii. BOVINE β-Bp CRYSTALLIN

Ac-A-S-N-H-E-T-Q-A-G-K-P-Q-P-L-H-P-K-I-I-I-F-E-Q-E-H-F-H-G-H-S-Q-E-L-N-P-G-N-P-C-L-K-E-T
-G-V-E-K-A-G-S-V-L-V-Q-A-G-P-W-V-G-Y-E-Q-A-N-C-K-G-E-Q-F-V-F-E-K-G-E-Y-P-R-W-D-S-W-T-S-S
-R-R-T-D-S-L-S-S-L-R-P-I-K-V-D-S-Q-E-H-K-I-T-L-Y-E-N-P-N-F-T-G-K-K-M-E-V-I-D-D-D-V-P-S-F
-H-A-H-G-Y-Q-E-K-V-S-S-V-R-V-Q-S-G-T-W-V-G-Y-Q-Y-P-G-Y-R-G-L-Q-Y-L-L-E-K-G-D-Y-K-D-S-G-D
-F-G-A-P-Q-P-Q-V-Q-S-V-R-R-I-R-D-M-Q-W-H-Q-R-G-A-F-H-P-S-S

Sequence from : Driessen, P.C., Herbrink, P., Bloemendel, H., deJong, W. (1981) J. Biochem. 121 , 83 - 91.

ix. PORCINE PEPSIN

I-G-D-E-P-L-E-N-Y-L-D-T-E-Y-F-G-T-I-G-I-G-T-P-A-Q-D-F-T-V-I-F-D-T-G-S-S-N-L-W-V-P-S-V-Y-
C-S-S-L-A-C-S-D-H-D-Q-F-N-P-D-D-S-S-T-F-E-A-T-S-Q-E-L-S-I-T-Y--G-T-G-S-M-T-G-I-L-G-Y-D-T
-V-Q-V-G-G-I-S-D-T-N-Q-I-F-G-L-S-E-T-E-P-G-S-F-L-Y-Y-A-P-F-D-G-I-L-G-L-A-Y-P-S-I-S-A-S-G
-A-T-P-V-F-D-N-L-W-D-Q-G-L-V-S-Q-D-L-F-S-V-Y-L-S-S-N-D-D-S-G-S-V-V-L-L-G-G-I-D-S-S-Y-Y-T
-G-S-L-N-W-V-P-V-S-V-E-G-Y-W-Q-I-T-L-O-D-S-I-T-M-D-G-E-T-I-A-N-I-Q-S-D-I-G-A-S-E-N-S-D-G
-E-M-V-I-S-C-S-S-I-D-S-L-P-D-I-V-F-T-I-N-G-V-Q-Y-P-L-S-P-S-A-Y-I-L-Q-D-D-D-S-C-T-S-G-F-E
-G-M-D-V-P-T-S-S-G-E-L-W-I-L-G-D-V-F-I-R-Q-Y-Y-T-V-F-D-R-A-N-K-V-G-L-A-P-V-A

Sequence from : Tang, J., Sepulveda, P., Marciniszyn, J., Chen, K.C.S., Huang, W.-Y., Tao, N., Liu, D., Lanier, J.P. (1973) Proc. Natl. Sci. U.S.A. 70 , 3437 - 3439.

PAGE MISSING IN ORIGINAL

x. BOVINE β-LACTOGLOBULIN

L-I-V-T-Q-T-M-K-G-L-D-I-Q-K-V-A-G-T-W-Y-S-L-A-M-A-A-S-D-I-S-L-L-D-A-Q-S-A-P-L-R-V-Y-V-E-
E-L-K-P-T-P-E-G-D-L-E-I-L-L-Q-K-W-E-N-(D/G)E-C-A-Q-K-K-I-I-A-E-K-T-K-I-P-A-V-F-K-L-D-A-I
-N-E-N-K-V-L-V-L-D-T-D-Y-K-K-Y-L-L-F-C-M-E-N-S-A-E-P-E-Q-S-L-(V/A)-C-Q-C-L-V-R-T-P-E-V-D
-D-E-A-L-E-K-F-D-K-A-L-K-A-L-P-M-H-I-R-L-S-F-N-P-T-L-Q-E-E-Q-C-H-I

Sequence from : Braunitzer, G., Chen, R., Schrank, B., Stang, A. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354 , 867 - 878.

xi. MELITTIN

G-I-K-A-V-L-K-V-L-T-T-G-L-P-A-L-I-S-W-I-K-R-K-R-Q-Q

Sequence from : Habermann, E., (1980) In: *Natural toxins* (Eaker, D. and Wadstrom, T. eds).

POOR SUBSTRATES (< 1.0 mol amine/mol protein incorporated)

xii. BOVINE PANCREATIC RIBONUCLEASE A

K-E-T-A-A-A-K-F-E-R-Q-H-M-D-S-S-T-S-A-A-S-S-S-N-Y-C-N-Q-M-M-K-S-R-N-L-T-K-D-R-C-K-P-V-N-T-F-V-H-E-S-L-A-D-V-Q-A-V-C-S-Q-K-N-V-A-C-K-N-G-Q-T-N-C-Y-Q-S-Y-S-T-M-S-I-T-D-C-R-E-T-G-S-S-K-Y-P-N-C-A-Y-K-T-T-Q-A-N-K-H-I-I-V-A-C-E-G-N-P-Y-V-P-V-H-F-D-A-S-V

Revised sequence from: Smyth, D.G., Stein, W.H., and Moore, H. (1963) J. Biol. Chem., 238 , 227-234.

Note: Disulphide bridges are between residues Cys_{26} and Cys_{84}, Cys_{40} and Cys_{95}, Cys_{58}, and Cys_{110}, and Cys_{65} and Cys_{72}.

xiii. HEN EGG LYSOSYME

K-V-F-G-R-C-E-L-A-A-A-M-K-R-H-G-L-D-N-Y-R-G-Y-S-L-G-N-W-V-C-A-A-K-F-E-S-N-F-N-T-Q-A-T-N-R-N-T-D-G-S-T-D-Y-G-I-L-Q-I-N-S-R-W-W-C-D-N-G-R-T-P-G-S-R-N-L-C-N-I-P-C-S-A-L-L-S-S-D-I-T-A-S-V-N-C-A-K-K-I-V-S-D-G-D-G-M-N-A-W-V-A-W-R-N-R-C-K-G-T-D-V-Q-A-W-I-R-G-C-R-L

Sequence from: Canfield, R.E. (1963) J. Biol. Chem. 238, 2698-2707; and Canfield, R.E. (1965) J. Biol. Chem. 240, 1997-2002.

Note: Disulphide bridges between Cys_{6} and Cys_{127}, Cys_{30} and Cys_{115}, Cys_{64} and Cys_{80}, and Cys_{76} and Cys_{94}.

xiv. HORSE HEART CYTOCHROME C

Ac-G-D-V-E-K-G-K-I-F-V-Q-K-C-A-Q-C-H-T-V-E-K-G-G-K-H-K-T-G-P-N-L-H-G-L-F-G-R-K-T-G-Q-A-P
-G-F-T-Y-T-D-A-N-K-N-K-G-I-T-W-K-E-E-T-L-M-E-Y-L-E-N-P-K-K-Y-I-P-G-T-K-M-I-F-A-G-I-K-K-K
-T-E-R-E-D-L-I-A-Y-L-K-K-A-T-N-E

Sequence from : Okunuki, K., Kamen, M.D. *Structure and Function of Cytochromes* University park press, 1968.

REFERENCES

ALLEN, G. (1981) In Laboratory Techniques in Biochemistry and Molecular Biology (WORK, T.S. and BURDON, R.H., eds.), vol 9, pp 30-31, Elsevier/North Holland, Amsterdam .

AKRIGG, D., BLEASBY, A.J., DIX, N.I.M., FINDLAY, J.B.C., NORTH, A.T.C., PARRY-SMITH, D., WOOTON, J.C., BLUNDELL, T.L., GARDNER, S.P., HAYES, F., ISLAM, S., STERNBERG, M.J.E., THORNTON, J.M., TICKLE, I.J., and MURRAY-RUST, P. (1988) Nature (London) 335 , 745 - 746.

AKROYD, P. (1965) Nature (London) 208 , 488 - 489

AMBRON, R.T., and KREMZNER, L.T., (1982) Proc. Natl. Sci. U.S.A. 79 , 3442.

ANDERSON, B.L., BERRY, R.W., TELSER, A. , (1983) Anal. Biochem. 132 , 365-375.

(a) ANDO,Y., ADACHI, M., UMEDA, K., MATSURA, A., NONAKA, M., UTCHIO, R., TANAKA, H., MOTOKI, M., (1989) Agric. Biol. Chem. 53 , (10) 2513-2617.

(b) ANDO, Y. , IMAMURA, S. , OWADA, M.K. , KANNAGI, T., KANNAGI, R., (1989) Biochem. Biophys. Res. Comm. 163 (2) , 944 - 951.

ANDO, Y. , IMAMURA, S. , OWADA, M.K. , KANNAGI, R. (1991) J. Biol. Chem. 266 , 1101 - 1108 .

ANDREEVA, N.S., GUSTCHINA, A.E., FEDOROV, A.A., SHUTZKEVER, N.E., VOLNOVA, T.V., (1977) Adv. Exp. Med. and Biol. 95 , 23 - 32.

ANDREEVA, N.S., and GUSTCHINA, A.E. (1979) Biochem. Biophys. Res. Comm. 87 (1) , 32 -42.

BAIROCH, A., and BOECKMAN, B., (1991) Nucleic acids research 19 , 2247-2249.

BAX, B., LAPATTO, R., NALIANI, V., DRIESSEN, H., LINDLEY, P.F., MAHADEVAN, D., BLUNDELL, T.L. and SLINGSBY, C. (1990) Nature (London) 347 , 776 - 780

BECKER, J.W. , REEKE, G.N., WANG, J.L., CUNNINGHAM, B.A., and EDELMAN, G.M., (1975) 250 , (4) , 1513-1524.

BECKER, J.W. and REEKE, G.N. (1985) Proc. Natl. Sci. U.S.A. 82 , 4225 - 4229.

BERBERS, G.A.M, BENTLAGE, H.C.M., BRANS, A.M.M., BLOEMENDEL, H., and DE JONG, W.W. (1983) Eur. J. Biochem. 135 , 315 - 320.

BERGAMINI, C.M. , SIGNORINI, M. , POLLTRONIER, L. (1987) Biophysica. et Biochemica. Acta. 916 , 149 - 151.

BERGAMINI, C.M. (1988) FEBS Lett. 239 (2) , 255 - 258.

BERGER, A. and SCHECHTER, I. (1970) Phil. Trans. R. Soc. London 257, 249-264

BERGMEYER, H.U., (1955) Biochem. Z. 327 , 255.

BIRCKBICHLER, P.J. , ORR, G.R., CONWAY, E., PATTERSON, M.K. (1977) Cancer

Research 37 , 1340-1344 .

BIRCKBICHLER, P.J., ORR, G.R., CARTER, H.A. and PATTERSON (1977) Biochem. Biophys. Res. Comm. 78 (1) , 1 - 7.

BJERRUM, HAWKINS, GRIFFIN and LORAND [unpublished results, cited in LORAND, L. and CONRAD, S.M. (1984) Mol. and Cellular Biochemistry 58 , 9 - 35.

BJORKMAN, P.J. , SAPER, M.A. , SUMRAOUL, B. , BENNETT, W.S. , STROMINGER, J.L. , WILEY, D.C. (1987) Nature (London) 329 , 506 - 518.

BLOCK, W.Z. (1954) Physiol. Chem. 296 , 108.

BLUNDELL, T., and WOOD, S., (1982) Ann. Rev. Biochem. 51 , 123 - 154.

BOOTHE, R.L., and FOLK, J.E., (1989) J. Biol. Chem. 244 , 399.

BORSOOK, H., DEASY, E.L. , HAAGEN-SMIT, A.J., KEIGHLEY, G., and LOWY, P.H., J. (1950) Biol. Chem. 184 , 529.

BOWEN, D., LITTLECHILD, J.A., FOTHERGILL, J.E., WATSON, H.C., HALL, L. (1988) Biochem. J. 254 , 509 - 517.

BOWNESS, J.M., FOLK, J.E. and TIMPL, R. (1987) J. Biol. Chem., 262 , 1022 - 1024

BRANDT, A. , GLANVILLE, R.W. , HORLEIN, D. , BRUCKNER, P. , TIMPL, R. , FIETZEK, P.P. , KUHN, K. (1984) Biochem. J. 219 , 625 - 634.

BRAUNITZER, G., Chen, R., Scrank, B., Stang, A., (1973) Hoppe-Seyler's Z. Physiol. Chem. 354 , 867 - 878.

BRENNER, S.C. and WOLD, F. (1978) Biochemie et Biophysica Acta. 522 , 71 - 83.

BUNGAY, P.J., OWEN, R.A., COUTTS, I.C., GRIFFIN, M. (1986) Biochem. J. 235 , 269.

BUSHNALL, G.W. , LOUIE, G.V. , BRAYER, G.D. (1990) J. Mol. Biol. 214 , 585 - 595.

BUTLER, S. and LANDON, M. (1979) Biochem. Soc. Trans. 7 , 397 - 398.

CAMBELL, R.L. , PETSKO, G.A. Biochemistry (1987) 26 , 8579 - 8584.

CANELLAKIS, Z.N., BONDY, P.K., and INFANTE, A.A. (1985) Proc.Natl. Sci. U.S.A. 82 , 7613 - 7615.

CANFIELD, R.E., (1963) J. Biol. Chem., 238 , 2698 - 2707.

CANFIELD, R.E., (1965) J. Biol. Chem. , 240 , 1997 - 2002.

CARIELLO, L. , WILSON, J. , LORAND, L. (1984) Biochemistry. 23 , 6843-6850.

CARRELL, N.A., ERICKSON, H.P., and McDONAGH (1989) J. Biol. Chem. 264 , 551 - 556.

CHAKRAVARTY, R. and RICE, R.H. (1989) J. Biol. Chem. 264 , 625 - 629.

CHANG, C.T., WU, C.S.C., and YANG, C.T. (1978) Anal. Biochem. 91 , 13.

CHEN, R. and DOOLITTLE, R.F. (1971) Biochemistry 10 , 4486 - 4491.

CHEN, Y.H., YANG, J.T., and MARTINEZ, H.M. (1972) Biochemistry 11 , 4120.

CHEN, K.Y., (1984) Mol. Cell. Biochem. 69 , 161.

CHOU, P.Y., and FASMAN, G.D., (1974a) Biochemistry 13 (2) , 211 - 222.

CHOU, P.Y., and FASMAN, G.D., (1974b) Biochemistry 13 (2) , 222 - 245.

CHUNG, S.I. (1972) Ann. N.Y. Acad. Sci. 202 , 240.

CLARKE, D.D. , MYCEK, M.J. , NEIDLE, A. , WAELSCH, H. (1959) Arch. Biochem.. Biophys. 79 , 535 - 534.

CLARKE, D.D. , NEIDLE, A. , SARKAR, N.K. and WAELSCH, H. (1957) Arch. Biochem. Biophys. 71 , 277.

CONNELLAN, J.M., CHUNG, S.I., WHETZEL, N.K., BRADLEY, L.M. and FOLK, J.E. (1971) J. Biol. Chem. 246 , 1093 - 1098.

COOKE, R.D. and HOLBROOKE, J. (1974) Biochem. J. 141 , 71-78 .

CORDELLA-MIELE, E., MIELE, L. and MUKHERJEE, A.B. (1990) J. Biol. Chem. 265 , 17180 - 17188.

COULSON, A.F.W., COLLINS, J.F., and LYALL, A. (1987) Computer J. 30 , 420 - 424.

COUSSONS, P.J., KELLY, S.M., PRICE, N.C., JOHNSON, C.M., SMITH, B. and SAWYER, L. (1991) Biochem. J. 273 , 73 - 78.

COUSSONS, P.J., KELLY, S.M. and PRICE, N.C. (1990) Biochem. Soc. Trans. 18 , 672 - 673.

CREAMER, L.K., PARRY, D.A.D. and MALCOLM, G.N. (1983) Arch. Biochem. Biophys. 227 , 98 - 105.

CREIGHTON, T.E., (1978) Prog. Biophys. Mol. Biol. 33 , 231 - 297.

DANGL, J.L., WENSEL, T.G., MORRISON, S.L., STRYER, L., HERZENBERG, L.A., OI, V.T., (1988) EMBO J. 7 (7) , 1989 - 1994.

DAVIES, P.J.A. , DAVIES, D.R. , LEVITZKI, A. , MAXFELD, F.R. , MILHAUD, P. , WILLINGHAM, M.C. and PASTAN, I.H. (1980) Nature (London) 283 , 162 - 167.

DE HAAS, G.H, SCHARRENBURG, G.J.M., SLOTBOOM, A.J., (1987) Biochemistry 26 , 3402 - 3408.

DEKKER, N., PETERS, A.R., SLOTBOOM, A.J., BOELENS, R., KAPTEIN, R., and DE HAAS, G., (1991) Biochemistry 30 , 3135 - 3147.

DICKERSON, R.E., (1972) Sci. Amer., 226 (4) , 58 - 72.

DIJKSTRA, B.W., DRENTH, J., KALK, K.H., and VANDERMAELEN, P.J., (1978) J. Mol. Biol. 124 , 53 - 60.

DIJKSTRA, B.W. , RENETSEDER, R. , KALK, K.H. , HOI, W.G.J. , DRENTH, J. (1983) J. MOL. BIOL. 168 , 163 - 179.

DOOLITTLE, R.F., (1973) Adv. In Protein Chem. 27 , 1 - 109 .

DOOLITTLE, R.F. , WATT, K.W.K. , COTTRELL, B.A. , STRONG, D.D. , RILEY, M. (1979) Nature (London) 280 , 464 - 468 .

DRENTH, J., JASONIUS, J.N., KOEKOEK, R., SWEN, H.M., WOLTHERS, B.G., (1968) Nature (London) 218 , 929.

DRENTH, J., JASONIUS, J.N., KOEKOEK, R., WOLTHERS, B.G., (1971) The Enzymes , ed. BOYER, P.D., 3rd Edn, Vol III, p485.

DUFTON, M.J., EAKER, D., HIDER, R.C., (1983) Eur. J. Biochem. 137 , 537 - 544.

DUNTEN, R.L., COHEN,R.E., GREGORI, L., and CHAU, V. (1991) J. Biol. Chem. 266 , 3260 - 3267.

DUTTON, A. and SINGER, S.J. (1975) Proc. Natl. Sci. USA 72 , 2568 - 2571.

DVILANSKY, A., BRITTEN, A.F.H., and LOEWY A.G., (1970) Brit. J. Haematology 18 , 399 - 410.

ELIOPOULOS, E., GEDDES, A.J., BRETT, M., PAPPIN, D.J.C., and FINDLAY, J.B.C.,

(1982) Int. J. Microbiol. 4 , 263 - 268.

EVENTOFF, W., and GURSKAYA, G.V., (1975) J. Mol. Biol. 93 , 55 - 62.

FESUS, L., ERDEI, A., SANDOR, M., (1984) Mol. Immunol. 21 , 1161.

FESUS, L. , FALUS, A. , ERDE, A. , KoLOMAN, L. (1981) J. Cell Biol. 89 , 706-710.

FESUS, L. and THOMAZY, V. (1988) Advances in Exp. Med. and Biol. 231 , 119-129.

FLANDERS, P., HUNT, S., REDDAWAY, S., and PARKINSON, D., (1978) High Speed Computer and Algorithm Organization. , ed. KLUCK, D.J., LAWRIE, D.H., and SAMEH, A.H., Achademic Press. London (1978), pp 113 - 120.

FOLK, J.E. and COLE, P.W. (1965) J. Biol. Chem. 248 , 2951 - 2960.

FOLK, J.E. (1983) Advances in Enzymology. 54 , 1 - 56.

FOLK, J.E. and GROSS, M. (1971) J. Biol. Chem. 246 , 6683 - 6691.

FOLK, J.E. and CHUNG, S.I. (1973) Adv. Enzymol. Relat. Areas Mol. Biol. 88 , 109 - 191.

FOLK, J.E. and COLE, P.W. (1965) J. Biol. Chem. 240 , 2951 - 2960.

FOLK, J.E. and COLE, P.W. (1966) J. Biol. Chem. 241 , 5518 - 5525.

FOLK, J.E. and COLE, P.W. (1965) J. Biol. Chem. 240 , 2951 - 2960.

GEORGE, M.D. , VOLBERG, T.M. , FLOYD, E.E. , STEIN, J.P. , JETTEN, A.M.J.B. (1990) J. Biol. Chem. 265 , 11098 - 11104.

GETTINGS, P. , CUNNINGHAM, L.W. (1986) Biochemistry 25 , 5011 - 5017.

GOMIS, R., CASANOVAS, A., MALAISSE, W.J., (1989) Biosci. Rep. 9 (1), 55 - 62.

GORMAN, J.J. and FOLK, J.E. (1981) J. Biol. Chem. 256 , 2712 - 2715.

GORMAN, J.J. and FOLK, J.E. (1980) J. Biol. Chem. 255 , 419 - 427.

GORMAN, J.J. and FOLK, J.E. (1984) J. Biol. Chem. 259 , 9007 - 9010.

GRAHAM, H.C., WILLIAMS, R.J.P., LITTLECHILD, J.A., WATSON, H.C. (1991) Eur. J. Biochem. 196 (2), 261 - 269.

GREEN,H. (1979) Harvey Lect. 74 , 101 - 139.

GREENFIELD, N., DAVIDSON, B., and FASMAN, G.D. (1967) , Biochemistry 6, 1630.

GROSS, M., and FOLK, J.E. (1973) J. Biol. Chem. , 248 , 1301 - 1306.

GROSSOWICZ, N., WAINFAN, E., BOREK, E., and WAELSCH, H., (1950) J. Biol. Chem. , 187 , 111 - 125.

GROVES, M.L., HIPP, N.J. and McMEEKIN, T.L. (1951) J. Amer. Chem. Soc. 73 , 2790 - 2793.

GUILD, B.C. and STROMINGER, J. (1984) J. Biol. Chem. 259 , 9236 - 9240.

GARNIER,, J., OSGUTHORPE, D.J., and ROBSON, B., (1978) J. Mol. Biol. 120 , 97-120.

HAND, D. , ELLIOT, B.M. , GRIFFIN, M. (1990) Biochimica Biophysica Acta. 1033 , 57 - 64.

HARDING, H.W.T., and RODGERS, C.E., (1971) Biochemistry 10 , 624 - 630.

HERSKOVITZ, T.T., (1965) J. Biol. Chem. 240 , 628.

HILGENFELD, R. , LIESUM, A. , STORM, R. , METZNER, H.J. , KARGES, H.E. (1990) FEBS Letts. 265 (1,2) , 110 - 112.

HITZEMAN, R.A., HAGIE, F.E., HAYFLICK, J.S., CHEN, C.Y., SEEBERG, P.H., and DERYNEK, (1982) Nucleic Acids Res. 10 , 7791 - 7808.

HOHL, D., MEHREL,T., LICHTI, U., TURNER, M.L., ROOP, D.R., and STEINER, P.M. (1991) J. Biol. Chem. 266 (10) , 6626 - 6636.

HOLLOSI, M. , KAJTAR, M. (1977) FEBS Lett. 74 , (2) 185 - 189.

HOLT, C. and SAWYER, L. (1988) Protein Engineering 2 (4) , 251 - 259 .

HORLEIN, D., FIETZEK, P.P., WACHTER, E., LAPIERE, C.M. and KUHN, K. (1979) Eur. J. Biochem. 99 , 31 - 38.

ICHINOSE, A. , BOTTENUS, R.E. and DAVIE, E.W. (1990) J. Biol. Chem. 265 , 13411 - 13413.

ICHINOSE, A. , HENDRICKSON, L.E. , FUJIKAWA, K. and DAVIE, E.W. (1986) Biochemistry 25 , 6900 - 6906.

IKURA, K. , NASU, T. , YOKOTA, H. , SASAKI, R. , CHIBA, H. (1987) Agric. Biol. Chem. 51 , 957 - 961.

IKURA, K. , NASU. T., YOKOTA, H., TSUCHITA, Y., SASAKI, R., CHIBA, C. (1988) Biochemistry 27 , 2898 - 2905.

IKURA, K. , YOKOTA, H., SASAKI, R. , CHIBA, H. (1989) Biochemistry 28 , 2344-2348.

INADA, Y., KUROZUMI, T., and SHIBATA, K., (1961) Arch. Biochem. Biophys. 93 , 30 - 36.

IWANIJ, V. (1977) Eur. J. Biochem. 80 , 359 - 368.

JENG, M. , ENGLANDER, W. , ELOVE, G.A. , WAND, J. , RODER, H. (1990)

Biochemistry 29 , 10433 - 10437.

JOSEPH, D., PETSCO, G.A., KARPLUS, M., (1990) Science 249 , 1425 - 1428.

KABSCH, W., MANNERHERZ, H.G., SUCK, D., PAI, E.F. and HOLMES, K.C. (1990) Nature (London) 347 , 37 - 44.

KASARDA, D.D. (1980) Ann. Technol. Agric. 29 , 151 - 173.

KELLA, N.K.D. and KINSELLA, J.E. (1988) Biochem. J. 255 , 113 - 118.

KIM, H.C. , IDELERS, W.W. , KIM, I.G. , HAN, J.H. , CHUNG, S.J., STEINERT, M. (1990) J. Biol. Chem. 268 , 536 - 539.

KIMURA, S. , AOKI, N. (1986) J. Biol. Chem. 261 , 15591 - 15595.

KLEE, C.B., CROUCH, T.H., RICHMAN, P.G. (1980) Ann. Rev. Biochem. 49 , 489 - 515.

KNIGHT, C.R.L. , REES, R.C. , ELLIOT, B.M. , GRIFFIN, M. (1990) Biochemica Biophysica Acta. 1053 , 13 - 20.

KORN, E.D., (1982) Physiol. Rev., 62 , 672 - 737.

KYTE, J., DOOLITTLE, R.F., (1982) 157 , 105 - 132.

LAEMMLI, U.K. (1970) Nature (London) 227 , 680 - 685.

LAZAROWITZ, S.G. , GOLDBERG, A.R. , CHOPPIN, P.W. (1973) Virology 56 , 172 - 180.

LEE, K.N. , BIRCKBICHLER, P.J. , PATTERSON, M.K. (1989) Biochem. Biophys. Res. Comm. 162 (3), 1370 - 1375 .

LI, C.H., (1964) Nature (London) 201 , 924.

LICHTARGE, O. , JARDETZKY, O., LI, C.H. (1987) Biochemistry 26 , 5916 - 5925.

LICHTI, V. , BEN, T. , YUSPA, S.H. (1985) J. Biol. Chem. 260 , 1422 - 1426.

LIM, V.I., (1974) J. Mol. Biol. 88 , 873 - 894.

LIN, W., and KASAMATSU, H., (1983) Anal. Biochem. , 112 302.

LIPMANN, F., and TUTTLE, C., (1954) J. Biol. Chem. 159 , 21.

LOEWY, A.G., WILSON, F.J., TAGGART, N.M.,GREENE, E., FRASCA, P., KAUFMAN, H.S., and SORRELL, M.J., (1983) Cell Motility 3 , 463.

LOEWY, A.G., MATACIC, S.S., (1981) Biochimica Biophysica. Acta. 668 , 167 - 176.

LORAND, L. , HAV, L.K.H. , SIEFRING, G.E. and RAFFERTY, N.S. (1981) Proc. Natl. Acad. Sci. USA 78 , 1356.

LORAND, L. (1988) Adv. Exp. Med. Biol. 231 , 79 - 94.

LORAND, L. and CAMPBELL, L.K. (1971) Anal. Biochem. 44 , 207 - 220 .

LORAND, L., CREDO, R.B. and JANUS, T.J. (1981) Methods in Enzymology 80 , 333 - 341.

LORAND, L. and CONRAD, S.M. (1984) Mol. and Cell Biol. 58 , 9 - 35 .

LYALL, A., HILL, C., COLLINS, J.F., and COULSON, A.F.W., (1986) Parallel Computing 85 (eds. Feilmeier, M., Joubert, G., and Schendel, U.) , Elsevier , Amsterdam, pp 235 - 240.

LYSTER, R.L.J. (1972) J. Dairy Res. 39 , 279 - 318

MACCIONI, R.B., SEEDS, N.W., (1986) Mol. Cell. Biochem. 69 , 161.

MANS, R.J. and NOVELLI, G.D. (1961) Arch. Biochem. Biophys. 94 , 48 - 53.

MARGOSIAK, S.A. , DHARMA, A. , BRUCE-CARVER, M.R. , GONZALES, A.P. , LOUIE, D. and KHEHN, G.D. (1990) Plant Physiol. 92 , 88 - 96.

MATSUDAIRA, P. (1987) J. Biol. Chem. 262 , 10035.

MAYO, K.H., BURKE, C. (1990) Biochemistry 29 , 3277 - 3286.

McDONAGH, R.P. , McDONAGH, J. , PETERSON, T.E. , THOGERSON, H.C. , SKORSTENGAARD, K. , SOTTRUP-JENSEN, L. , MAGNUSSON, S. (1981) FEBS Lett. 127 (2) , 174 - 178.

McDOWELL, J.H. , UBEL, A. , BROWN, R.A. , HARGREAVES, P.A. (1986) Arch.

Biochem. Biophys. 249 (2) , 506 - 514.

McKENZIE, H.A. and SAWYER, W.H. (1967) Nature (London) 214 , 1101 - 1105.

MEHREL, T. HOHL, D. , ROTHNAGEL, J.A. , LONGLEY, M.A. , BUNDMAN, D. , CHENG, C. , LIHTI, V. , BISHER, M.E. , STEVEN, A.C. , STEINERT, P.M. , YUSPA, S.H. , ROOP, D.R. (1990) Cell 61 , 1103 - 1112.

MEHTA, K. , RAO, U.R. , VICKERY, A.C. and BIRCKBICHLER, P.J. (1990) Biochem. Biophys. Res. Comm. 173 (3) , 1051 - 1057.

MENNE, C.,SUSKE, G., ARNEMANN, J., WENZ, M., CATO, A.C.B., and BEATO, M., (1982) Proc. Natl. Sci. U.S.A. 79 , 4853 - 4857.

METAFORA, S. , FACCHIANO, F. , FACCHIANO, A. , ESPOSITO, C. , PELUSO, G. , PORTA, R. (1987) J. Protein Chem. 6 (4) , 353 - 359.

MILES, M.J., CARR, H.J., McMASTER, T.C., I'ANSON, K.J., BELTON, P.S., MORRIS, V.J., FIELD, J.M., SHREWRY, P.R., TATHAM, A.S., (1991) Proc. Natl. Sci. U.S.A. 88 , 68 - 71.

MORIZE, I. , SURCOUF, E. , VANEY, M.C. , EPELBOIN, Y. , BUEHNER, M. , FRINDLANSKY, F. , MILGROM, E. , MORNON, J.P. (1987) 194 , 725 - 739.

MORTENSEN, S.B. , SOTTRUP-JENSEN, L. , HANSEN, H.F. , RIDER, D., PETERSEN, T., MAGNUSSON, S. (1981) FEBS Letts 129 (2) 314 -317.

PETERSEN, T.E., MAGNUSSON, S. (1981) FEBS Lett. 129 , (2) 314 - 317.

MOSHER, D.F. (1976) J. Biol. Chem. 251 , 1639 - 1645 .

MURTHY, R.N., REID, T.J., SICIGNANO, A., TANAKA, N., ROSSMANN, M.J. (1981) J. Mol. Biol. 152 , 465 - 499.

MYCEK,M.J., CLARKE, D.D., NEIDLE, A., and WAELSCH, H. (1959) Arch. Biochem. Biophys. 84 , 528.

NEIDLE, A. and ACS, G. (1961) Fed. Proc. Am. Soc. Exp. Biol. 20 (1), 234.

NOMOTO, M., NARAHASHI, Y., and MURAKAMI, M., (1960) J. Biochem. (Tokyo) 48 , 593.

NORTHROP, J.H., KUNITZ,M., and HERRIOT, R.M. (1948) Crystalline Enzymes 2nd Edition, New York, N.Y., Columbia Univ. Press, p120.

NURY, S. and MEUNIER, J.C. (1990) Biochemie 72 , 219 - 226.

OVCHINNIKOV, Y.A., (1987) T.I.B.S. 12 , 434 - 438.

PAPIZ, M.Z., SAWYER, L., ELIOPOULOS, E.E., NORTH, A.C.T., FINDLAY, J.B.C., SIVAPRASADORAO, R., JONES, T.A., NEWCOMER, M.E. and KRAULIS, P.J. (1986) Nature (London) 324 , 383 - 385.

PARAMESWARAN, K.N. , VELASCO, P.T. , WILSON, J. , LORAND, L. (1990) Proc. Natl. Sci. U.S.A. 87 , 8472 - 8475.

PEARSON, and LIPMAN (1987) Methods in Enzymology 183 , 146 -159.

PEPINSKY, R.B., SINCLAIR, L.K., CHOW, E.P. and O'BRINE-GRECO, B. (1989) Biochem. J. 263 , 97 - 103.

Biochem. J. 263 , 97 - 103.

PEREZ-PAYA, E., THIAUDIERE, E., ABAD, C., DUFOURCQ, J. (1991) FEBS Letts. 278 (1) , 51 - 54.

PERKINS, R.E., CONROY, S.C., DUNBAR, B., FOTHERGILL, L.A., TUITE, M.F., DOBSON, M.J., KINGSMAN, S.M. and KINGSMAN, A.J. (1983) Biochem. J. 211 , 199 - 218.

PHILLIPS, D.C., (1967) , Proc. Natl. Sci. U.S.A. 57 , 484.

PHILLIPS, M.A. , STEWART, B.E. , QIN, Q. CHACKRAVARTY, R., FLOYD, K.E. , JETTEN, A.M. , RICE, R.H. (1990) Proc. Natl. Sci. U.S.A. 87 , 9333 - 9337 .

PINCUS, J.H., WAELSCH, H. (1968) Arch. Biochem. Biophys. 126 , 34 - 43.

PLACENTINI, M. and BENINATI, S. (1988) Biochem. J. 249 , 813 - 817.

POBER, J.S., IWANIJ, V., REICH, E., STRYER, L. (1978) Biochemistry 17 , 2163 - 2169.

POBER, J.S. and STROMINGER, J.L. (1981) Nature (London) 289 , 819 - 821.

PORTA, R., ESPOSITO, C., METAFORA, S., MALORNI, A., PUCCI, P., SICILIANO, R., and MARINO, G. (1991) Biochemistry 30 , 3114 - 3120.

PORTA, R., ESPOSITO, C., GENTILE, V., MARINIELLO, L., PELUSO, G., and METAFORA, S. (1990) Int. J. Peptide Protein Res. 35 , 117 - 122.

PORTA, R., ESPOSITO, C, METAFORA, S., MALORNI, A., PUCCI, P. and MARINO, G. (1989) Adv. Exp. Med. Biol. 250 , 403 - 409.

PORTA, R., GENTILE, V., ESPOSITO, C., FUSCO, A., PELUSO, G., METAFORA, S. (1988) Neuropeptides 11 , 89 - 92.

PORTA, R., GENTILE, V., ESPOSITO, C., MARINIELLO, L., AURICCHIO, S., (1990) Phytochemistry 29 (9), 2801 - 2804.

PROVENCHER, S.W. and Glöckner, J. (1981) Biochemistry 20 , 33 - 37.

PUCCI, P., MALORNI, A., MARINO, G., METAFORA, S., ESPOSITO, C., and PORTA, R. (1988) Biochemical Biophysical Res. Comm. 154 (2) , 735 - 740.

PUCCI, P., MALORNI, A., MARINO, G., METAFORA, S., ESPOSITO, C. and PORTA, R. (1988 b) Adv. Exp. Med. Biol. 231 , 73 - 78.

REEKE JN.,G.N., BECKER, J.W. EDELMAN, G.M., J. Biol. Chem. (1975) 250 (4) , 1525 - 1547.

RICE, R.H. , GREEN, H. (1977) Cell 11 , 417 - 422.

RUSSELL, P., ZELENKA, P., MARTENSEN, T., REID, T.W. (1987) Curr. Eye. Res. 6 , 533.

RUSSELL, D.H. (1981) Medical Biology 59 , 286 - 295.

SASAKI, K., DOCKERILL, S., ADIAMIAK, D.A., TICKLE, I.J. and BLUNDELL, T.L. (1975) Nature (London) 257 , 751 - 757.

SASKI, K., DOCKERILL, S., ADAMIAK, D.A., TICKLE, I.J., BLUNDELL, T.L., (1975) Nature (London) 257 , 751 - 757.

SAWYER, L., FOTHERGILL-GILMORE, L.A., and FREEMONT, P.S. (1988) Biochem. J. 249 , 789 - 793.

SAWYER, L., FOTHERGILL-GILMORE, L.A., and RUSSELL, G.A., (1986) Biochem. J. 236 127 - 130.

SAWYER, L., PAPIZ, M.Z., NORTH, A.C.T. and ELIOPOULOS, E.E. (1985) Biochem. Soc. Trans. 13 , 265 - 266.

SAWYER, L., TOLLIN, P., and WILSON, H.R., (1987) J. Gen. Virol. 68 , 1229 - 1233.

SAXENA, V.P., and WETLAUFER, D.B. (1971) Proc Natl. Sci. U.S.A. 68 , 969.

SCHROEDER, W.A., SHELTON, J.R., SHELTON, J.B., ROBERTSON, B., APELL, G., FANG, R.S., BONAVENTURA, J. (1982) Arch. Biochem. Biophys. 214 , 397 - 421.

SCHROEDER, W.A., SHELTON, J.R., SHELTON, J.B., OLSON, B.M. (1964) Biochim. Biophys. Acta. 89 , 47 - 65.

SEDMAK, J.J. and GROSSBERG, S.E. (1977) Anal. Biochem. 79 , 544 - 552.

SELKOE, D.J., ABRAHAM, C., IHARA, Y. (1982) Proc. Natl. Sci. U.S.A. 79 , 6070 -

6074.

SICHAK, S.P., DOUNCE, A.L. (1987) Biochimica et Biophysica Acta. 925 , 282 - 289.

SIEGEL, J.R., STEINMETZ, W.E. and LONG, G.L. (1980) Anal. Biochem. 104 , 160 - 167.

SIMON, M. and GREEN, H. (1985) Cell. 40 , 677 - 683.

SIMON, M. and GREEN, H. (1988) J. Biol. Chem. 263 , 18093 - 18098.

SKORSTENGAARD, K., HALKIER, T., HOJRUP, P., MOSHER, D. (1990) FEBS Letts. 262 (2) , 269 - 274.

SMITH, T.F., and WATERMAN, M.S., (1981) J. Mol. Biol. 147 , 195 - 197.

SPRINGER, T.A., and STROMINGER, J.L. (1976) Proc. Natl. Sci. U.S.A. 73 (7) , 2481 - 2485).

SMYTH, D.G., STEIN, W.H., MOORE, S. (1963) , 238 , 277 - 234.

SPRAGG, S.P., WILCOX, J.K., ROCHE, J.J., BARNETT, W.A., (1976) Biochem. J. 153 , 423 - 428.

STEIN, P.E., LESLIE, A.G.W., FINCH, J.T., TURNELL,W.G., McLAUGHLIN, P.J., CARRELL, R.W., (1990) Nature (London) 347 , 99 - 102.

STERN, R.A., TROJANOWSKI, J.Q., LEE, V.M.Y., (1990) FEBS Letts. 264 (1) , 43 - 47.

SUNG, L.A., CHIEN, S., CHANG, L.S., LAMBERT, K., BLISS, S.A., BOUHASSIRA, E.E., NAGEL, R.S., RYBICK, A.C. (1990) Proc. Natl. Sci. U.S.A. 87 , 955 - 959.

SYGUSCH, J., BEAUDRY, D., ALLAIRE, M., (1987) Proc. Natl. Sci. U.S.A. 84 7846 - 7850.

TABOR, C.W., and TABOR, H., (1984) Ann. Rev. Biochem. 53 , 749.

TAKAGI, T. and DOOLITTLE, R.F. (1974) Biochemistry 13 , 750 - 756 .

TAKAHASHI, N., TAKAHASHI, Y. and PUTNAM, F.W. (1986) Proc. Natl. Acad. Sci. U.S.A. 83 , 8019 - 8023.

TAKASHI, R., (1988) Biochemistry 27 , 938 - 943.

TAKEDA, K. and MORORIYAMA, Y. (1989) J. Prot. Chem. 8, 487 - 494.

TANFORD, C. and TAGGART, V.G. (1961) J. Amer. Chem. Soc. 83 , 1634 - 1638.

TANG, J., SEPOVELDA, P., MARCINISZYN, J., CHEN, K.Y.S., HUANG, W.Y., TOO, N., LANIER, J.P. (1973) Proc. Natl. Sci. U.S.A. 70 , 3437.

TATHAM, A.S., MARSH, N.N., WEISNER, H., SHREWRYS, P.R. (1990) Biochem. J. 270 , 313 - 318.

TERWILLIGER, T.C., and EISENBERG, D.J. (1982) J. Biol. Chem. 257 , 6061 - 6022.

THEORELL, H. (1951) The Enzymes (Sumner, J.B., and Myrback, K., eds.) pp 397 -

427, Achademic Press New York.

THOMPSON, M.P.. and PEPPER, L., (1964), J. Dairy Sci. 47 , 633.

TODA, H., FOLK, J.E., (1969) Biochimica. et Biophysica. Acta. 175 , 427 - 430.

TOLAN, D.R., AMSDEN, A.B., PUTNEY, S.D., URDEA, M.S., and PENHOET, E.E., (1984) J. Biol. Chem. 259 , 1127 - 1131.

UDGAONKAR, J.B. and BALDWIN, R.L. (1990) Proc. Natl. Sci. U.S.A. 87 , 8197 - 8201.

URRY, D.W., VENKATACHALAN, C.M., KHALED, M.A., and PRASAD, K.V. (1981) Neurohypophyseal hormones and other biologically active peptides; Developments in Endocrinology 13 David H. Sclesinger (ed.) Elsevier/North Holland press.

VEGA, M.A. and STROMINGER, J.L. (1989) Proc. Natl. Sci. U.S.A. 86 , 2688 - 2692.

VON LEUVEN, F. (1982) T.I.B.S. 185 - 187.

WATSON, H.C., WALKER, N.P.C., SHAW, P.J., BRYANT, T.N. WENDELL, P.L., FOTHERGILL, L.A., PERKINS, R.E., CONROY, S.C., DOBSON, M.J., TUITE, M.F., KINGSMAN, A.J. and KINGSMAN, S.M. (1982), EMBO. J. 1 , 1635 - 40.

WEBSTER, R.G. , LAUER, W.G., AIR, G.M. , SCHILD, G.C. (1982) Nature (London) 296 , 115 - 121.

WHITE, M.F., and FOTHERGILL-GILMORE, L.A. (1988) FEBS Letts. 229 (2) , 383 -

387.

WILLIAMS, R.W. , WEAVER, J.L. (1990) J. Biol. Chem. 265 , 2505 - 2513.

WILSON, C.A.B., HARDMAN, N.. FOTHERGILL-GILMORE, L.A., GAMBLIN, S.J. and WATSON, H.C. (1987) Biochem. J. 241 , 609 - 614.

WILSON, I.A. , SKEHEL, J.J. , WILEY, D.C. (1981) Nature (London) 289 , 366 - 375.

WINN, S.I., WATSON, H.C., HARKINS, R.N. and FOTHERGILL, L.A. (1981) Phil. Trans. R. Soc. London 3, 121 - 130.

WINTER, G.W., and MILSTEIN, C. (1991) Nature (London) 349 , 293 - 299.

WISTOW, G., SLINGSBY, C., BLUNDELL, T., DRIESSEN, H., DE JONG, W. and BLOEMENDAL, H. (1981) FEBS Letts. 133 , 9 - 16.

WLODAWER, A., and SJOLIN, L., (1983) Biochemistry 22 , 2720 - 2728.

WOLD, F. (1985) Trends Biochem. Sci. 10 , 4 - 6.

WONG, W.S.D., BATT, C., KINSELLA, J.E. (1990) Int. J. Biochem. 22 (1) , 53 - 59.

YAMANISHI, K., LIEW, F., KONISHI, K., YASUNO, H., DOI, H., HIRANO, J., FUKUSHIMA, S. (1991) Bioche. Biophys. Res. Comm. 175 (3) , 906 - 913.

YAN, B. and WOLD, F. (1984) Fed. Proc. Am. Soc. Exp. Biol. 43 , 1783.

YAN, B., WOLD, F. (1984) Biochemistry 23 , 3759 - 3765.

YANG, J.T. , BEWLEY, T.A. , CHEN, G.C. , LI, C.H. (1977) Proc. Natl. Sci. U.S.A. 74 , 3235 - 3238.

YANG, J.T., and SAMEJIMA, T., (1963) J. Biol. Chem. 238 , 3262 - 3267.

YOSHIMURA, M., BANERJEE, M.R., OKA, T., (1986) Nucleic Acid Res. 14 , 8224.

ZAPPIA, V., GALLETTI, P., PORTA, P., and WOLD, F. (eds). (1988) Adv. Exp. Med. Biol. 231.

ZATIOUKAL, K., DENK, H., LACKINGER, E., RAINER, I. (1989) Lab Invest. 61 , 603.