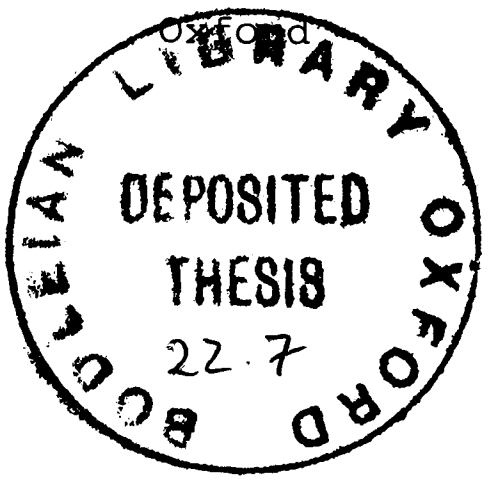
STRUCTURE AND ACTIVITY OF FACTOR $\bar{D}$ OF THE ALTERNATIVE PATHWAY OF HUMAN COMPLEMENT

Diana Margaret Ashley Johnson

Wolfson College

Oxford



A Thesis Submitted for the

Degree of Doctor of Philosophy

MRC Immunochemistry Unit

Department of Biochemistry

University of Oxford

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Ken Reid, for his help and advice throughout the course of this project. I am also grateful to all the members of the MRC Immunochemistry Unit and in particular, to Dr. Bob Sim, Dr. Edith Sim and Dr. Jean Gagnon, for many helpful discussions. I am indebted to Tony Gasgoyne for performing endless amino acid analyses, and to Dr. Jean Gagnon and Tony Willis whose expertise in automated sequencing was invaluable. My thanks are also extended to Alice Moseby and Sue Saunders for their typing skills, and Frank Caddick for photography and diagrams. The financial assistance of the British Council and the Association of Commonwealth Universities who provided a Commonwealth Scholarship is gratefully acknowledged, as is the kindness of Prof. R.R. Porter in making available the facilities of the Unit. Finally, I owe an incalculable debt to my family and friends for their support and encouragement.

ABBREVIATIONS

AGLME	acetyl-glycyl-L-lysine methyl ester
ALME	N-α-acetyl-L-lysine methyl ester
AT-III	antithrombin-III
BAEE	N-benzoyl-L-arginine ethyl ester
BDC-OH	4,4'-bis-dimethyldiphenylcarbinol
B_z	benzoyl
CT	chymotrypsin
CTG	chymotrypsinogen
CVF	cobra venom factor
Dansyl	1-dimethylnaphthalene-5-sulphonyl
DNP	2,4-dinitrophenol
EDTA	ethylene-diamine-tetraacetic acid, disodium salt
EGTA	ethyleneglycol-bis-(β-amino-ethyl ether)-N,N'-tetraacetic acid
HPLC	high pressure liquid chromatography
IAA	iodoacetic acid
IBA	iodosobenzoic acid
LBTI	limabean trypsin inhibitor
pNPGB	p-nitrophenyl-p'-guanidinobenzoate
PAS	periodic acid/Schiff
PBS	sodium phosphate buffer (at the stated molarity and pH) containing 0.15 M NaCl
PMSF	phenylmethylsulphonyl fluoride
PTH	phenylthiohydantoin
iPr_2P-F	diisopropylphosphofluoridate
R(D)	factor $\bar{D}$-deficient serum
SBTI	soybean trypsin inhibitor
SCM	S-carboxylmethyl
SDS	sodium dodecylsulphate

TAME	N-tosyl-L-arginine methyl ester
TPCK	tosylamide-2-phenylethyl-chloromethyl ketone
ZLSB	thiobenzyl-benzoxycarbonyl-lysinate

Other abbreviations conform to those recommended by the Biochemical Society (Biochem. J. (1975) 145, 1-23). The one letter code for amino acids is from Biochem. J. 113, 1-4.

NOMENCLATURE

The nomenclature for classical complement components conforms to that described in Bull. W.H.O. (1968) 30, 395. For the alternative complement pathway, it is that provisionally adopted by the Complement Nomenclature Committee, Brighton, 1974.

Sheep erythrocytes (E) coated with rabbit antibody to sheep erythrocyte stromata (A) are designated EA followed by the name of complement components which are bound on the cell surface.

Nomenclature for immunoglobulin and immunoglobulin fragments is standard.

ABSTRACT

Structure and activity of factor $\bar{D}$ of the alternative pathway of human complement

Diana Margaret Ashley Johnson

Wolfson College, Oxford

Submitted for the degree of D. Phil. Michaelmas Term, 1980

1. A method for the purification of the serine protease, factor $\bar{D}$, was developed using conventional chromatographic procedures. The final product was homogeneous as judged by SDS/polyacrylamide gel electrophoresis, its migration as a single component in ion exchange and gel filtration media, and its amino acid sequence analysis. The molecule had an apparent molecular weight of 24,000. It contained <1.5% (w/w) reducing sugars as judged by periodic acid/Schiff staining, and existed as a monomer in buffers containing either EDTA or calcium ions.

2. Approximately 84% of the amino acid sequence was established unequivocally by automated sequence analysis of the intact molecule and peptides derived by digestion with CNBr, o-iodosobenzoic acid, trypsin and V8 protease. Carboxypeptidase-Y digestion was used to establish the C-terminal amino acid. The peptides were aligned either by homology with other serine proteases, or by the overlap of sequences obtained from peptides derived by different fragmentation procedures. The molecule had a typical serine protease-type sequence with isoleucine as the N-terminal amino acid. The active site serine and aspartic acid and the surrounding sequences were conserved as well as the sequence around the position of the active site histidine, although this residue itself was not identified.

3. The possibility of the existence of a factor $\bar{D}$ zymogen which can be activated by trypsin was reinvestigated, but no evidence for a precursor was found. No enzymic activity towards a number of p-nitroanilide substrates and arginyl and lysyl esters was observed with factor $\bar{D}$, but it was found to release p-nitrophenol from p-nitrophenyl-p'-guanidinobenzoate. Factor $\bar{D}$ was inhibited by diisopropylphosphofluoridate and p-nitrophenyl-p'-guanidinobenzoate, but a variety of other non-protein and protein inhibitors including α_2-macroglobulin, $C\bar{1}$ inhibitor and inter-α-trypsin inhibitor had no effect on enzymic activity.

PUBLICATIONS

1. D.M.A. Johnson, J. Gagnon and K.B.M. Reid (1980) "Factor $\bar{D}$ of the alternative pathway of human complement". Biochem. J. 187, 863-874.

   A copy of this publication is included at the end of the thesis.

2. K.B.M. Reid and D.M.A. Johnson. "Preparation of human factor $\bar{D}$ of the alternative pathway of complement". Methods in Enzymology (in press).

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

Factor $\bar{D}$ belongs to the group of endopeptidases known as serine proteases which rely for their activity on an unusually reactive serine residue located in the active site. Its physiological role is in the complement pathway, a system of about twenty proteins which occur in vertebrate blood plasma and which plays a major role in immune defence against microorganisms.

Complement acts as a cascade system, and it may be compared in this respect with other mechanisms of blood plasma such as clotting, fibrinolysis and the renin-angiotensin and kinin-forming systems. In all these processes, an initial stimulus triggers the first component of the cascade and the sequential activation of subsequent components results in a rapid and amplified response. In the case of complement, the stimulus may be an immune complex although antibody-independent activation may also occur. If the original target was a cell surface, the end-result can be cell lysis. Integral to the system are the control mechanisms required to prevent accidental or excessive stimulation which could endanger the organism. In complement this control is exercised both through the intrinsic decay rates of the intermediates and through the action of control proteins.

There are two distinct pathways of complement activation called for historical reasons, the classical pathway and the alternative pathway. These differ in their early components, but the central event of activation by either pathway is the cleavage of component C3 by a C3 convertase. C3 is a major plasma protein and is present in human plasma at a concentration of 1.2 mg/ml. Following C3 activation, the two pathways converge and utilise the same terminal components, C5-C9. The two pathways of complement activation are shown in Fig. 1.1 and some properties of the components are summarised in Table 1.1.

The alternative pathway can function independently of specific antibody and, in a simplified view, may form a first line of defence against infection (Reid and Porter, 1981). It is conceivable, therefore, that it is the more primitive of the two pathways in evolutionary terms, and arose from an "archeo-complement" system (Lachmann, 1979; Lachmann and Hobart, 1979). The classical pathway would then have arisen later to provide a response to antibody-mediated reactions. It is possible that the terminal components, C5-C9, also arose from the "archeo-complement" system, although the evidence for this is less convincing (Lachmann, 1979). These possible evolutionary relationships are of interest as particular components of the classical and alternative pathways do show strong structural and functional similarities. For instance, factor $\bar{D}$ is possibly similar to the catalytic chains of $C\bar{1}r$ and $C\bar{1}s$; factor B is similar to C2, and these two are similar in molecular weight and chain structure to C6 and C7

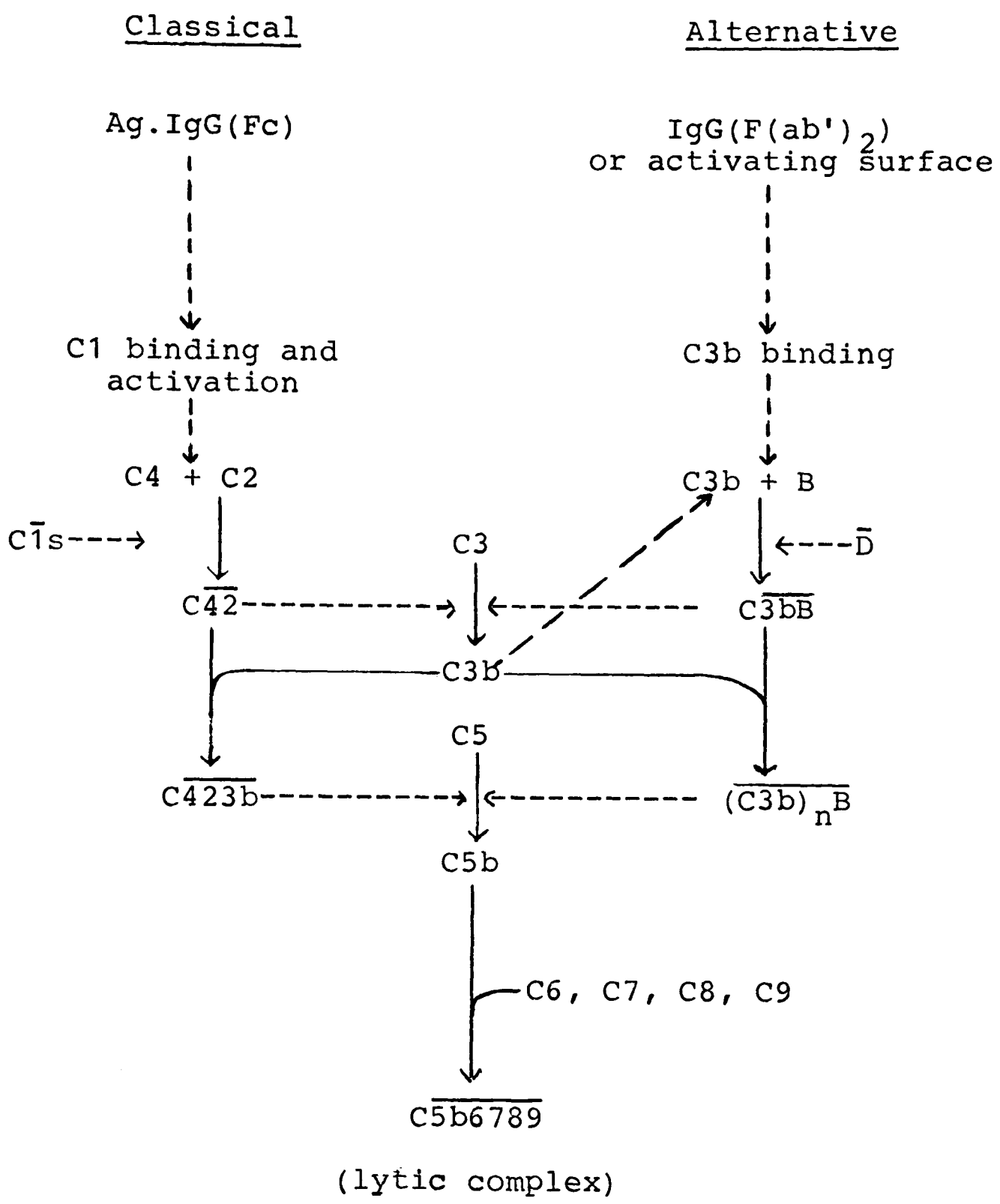


Fig. 1.1 Pathways of complement activation

Table 1.1 Components of the human complement system

Protein	Serum Concentration µg/ml	Sedimentation coefficient	Mol.† Wt.	Number of Chains
C1q	180	11.1	400,000	18
C1r	50-100	7.5	170,000	2
C1s	20-110	4.5	85,000	1
C4	400	10.0	206,000	3
C2	15-20	4.5	100,000	1
C3	1200	9.5	185,000	2
$\bar{D}$	1-2	3.0	24,000	1
C3b	trace	9.0	171,000	2
B	200	5.0-6.0	93,000	1
C5	80	8.7	190,000	2
C6	75	5.5	128,000	1
C7	55	6.0	121,000	1
C8	80	8.0	150,000	3
C9	230	4.5	79,000	1
Properdin	25	5.4	220,000	4
C3NeF	Pathological	7.0	150,000	4
$F\bar{42}$	Pathological	-	150,000	-
$C\bar{1}$ inhibitor	240	4.5	100,000	1
C3b inactivator	10-20	5.5-6.0	88,000	2
β1H	670	-	150,000	1
C4 binding protein	300	10.0-11.0	540-570,000	?8
S protein	-	-	70,000	1

†mol. wts. are for nondissociating conditions.

-, no data available.

(Lachmann and Hobart, 1979); C3, C4 and C5 also show similarities in gross structure, properties and control mechanisms.

$C\bar{1}r$, $C\bar{1}s$, factor $\bar{D}$, C3b INA (J. Gagnon and L.M. Hsiung, unpublished results), C2 and factor B are all serine proteases. The serine proteases are widely spread in nature and also include several blood-clotting proteins, pancreatic and microbial enzymes. They are usually secreted as zymogens which upon activation give rise to a catalytic unit of 20,000-30,000 mol. wt. In the case of the pancreatic proteases, the mol. wts. of the zymogens and activated enzymes are similar, whereas the zymogens of the coagulation and complement systems often exceed a mol. wt. of 60,000. The proteases of the complement system are also unusual in that, unlike pancreatic serine proteases, they have extremely rigid prerequisites and narrow specificities for cleaving polypeptide substrates; factor $\bar{D}$ and possibly C3b INA appear to lack a zymogen; and C2 and factor B have an unusual N-terminal structure. These enzymes represent at least three groups of interesting serine proteases whose structural and functional analysis should add greatly to the understanding of serine protease catalysis and evolution.

1.2 THE COMPLEMENT SYSTEM

The major established functions of complement are the production of membrane lesions, the production of biologically active peptides, and immune adherence to a variety of cell types (Lachmann, 1973). It has also been suggested that complement takes part in cell-mediated cytotoxicity reactions (Sundsmo *et al*, 1978a; 1978b) and that some components may play a role as cell-surface proteins (Bodmer, 1976; Barnstable, 1978).

1.2.1 The alternative pathway

Complement was first recognised in the late nineteenth century (von Fodor, 1886; Nuttal, 1888; Buchner, 1889; Bordet, 1896) and the existence of the alternative pathway is evident from early work on yeast (von Dungern, 1900; Whitehead, 1925) and cobra venom (Braun, 1911; Ritz, 1913). Although Pillemer *et al* (1953, 1954, 1955) had found evidence for an alternative pathway of complement activation in their well known experiments on zymosan-dependent utilisation of C3, it was only in the late 1960's that interest in the pathway escalated. It is now known to consist of three main components, factors B, $\bar{D}$ and C3, together with properdin and the control proteins C3b INA and β1H. The pathway is summarised in Fig. 1.2.

1.2.1.1 Factor $\bar{D}$

Factor $\bar{D}$ is the smallest protein of the complement pathway (mol. wt. 24,000) and is a serine protease, as is indicated by its inhibition by iPr_2P-F (Fearon *et al*, 1974). Its function in complement activation is to activate factor

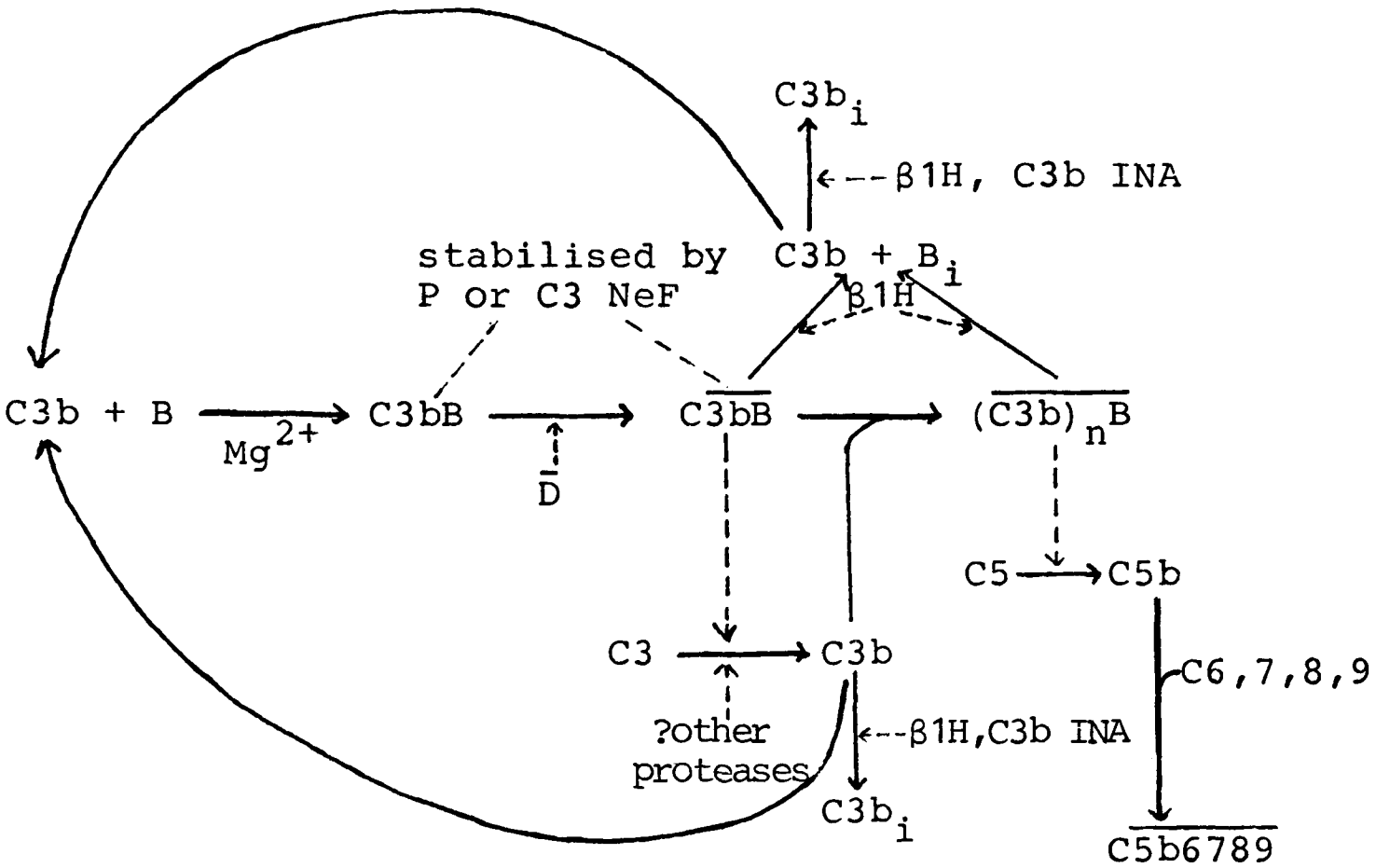


Fig. 1.2 The alternative pathway

C3b is continuously generated, and inactivated by β1H and C3b INA. In the presence of an activator, C3b becomes surface bound and protected from the regulatory proteins, so that amplification is possible.

B in the Mg^{2+}-dependent complex, C3bB, and so to form the active alternative pathway C3 convertase. The chemistry and activity of factor $\bar{D}$ are discussed more fully in section 1.4.

1.2.1.2 Factor B

Factor B is a glycoprotein of mol. wt. 93,000. It is cleaved in the C3bB complex by factor $\bar{D}$ to form two fragments, Ba (mol. wt. 30,000) and Bb (mol. wt. 63,000). The Ba fragment which does not participate further in the pathway, has been reported to have chemotactic activity (Hadding et al, 1978). When complexed to C3b, the larger Bb fragment can cleave C3 to C3b (Götze and Müller-Eberhard, 1971). Medicus et al (1976) have reported iPr_2P-F-inhibition of factor B and that iPr_2P-F was incorporated into the Bb fragment. Although this has not been corroborated by others (Kerr, 1979), amino-acid sequence analysis of factor B (Niemann et al, 1980; Christie et al, 1980) has shown unequivocally that Bb has the expected structural characteristics of a serine protease with the highly conserved active site residues (histidine, aspartic acid and serine) present in positions homologous to those of other typical serine proteases if comparisons are made from the C-terminal end (Fig. 6.1). The N-terminal of Bb is unlike other serine proteases, having an additional 300 amino acid residues. It is probable that these are involved in the interaction with C3b. This unusual N-terminal sequence suggests that the activation mechanism for factor B is unlike that of more typical serine proteases,

but probably similar to that of C2 which it resembles in gross structure (Kerr, 1979).

Besides participating in the C3 convertase, there have been reports that Bb can induce monocyte spreading (Sundsmo and Götze, 1980) and can bind DNA (Gardner et al, 1980).

The structural gene for factor B is closely linked to that of C2 and occurs on chromosome 6 in humans, very close to the HLA-B locus (Lachmann and Hobart, 1979).

1.2.1.3 C3

C3 cleavage is the central event of complement activation by either the classical or alternative pathways. C3 is a glycoprotein of mol. wt. 185,000 and is composed of two polypeptide chains, α (mol. wt. 105,000) and β (mol. wt. 75,000), linked by one or more disulphide bonds. C3 is synthesized as a single polypeptide chain (Brade et al, 1977b) with the α and β chains aligned in the order, β-α (Colten, 1980 On activation by a C3 convertase a fragment, 77 amino acids long, is released from the N-terminal of the α chain (C3a). C3a is not involved further in the complement sequence but has chemotactic (Goetzl and Austen, 1974) and anaphylatoxic properties (Dias da Silva and Lepow, 1967).

C3b can participate in several biological functions via a stable binding site. These include immune adherence reactions and phagocytosis. Nascent C3b also has a labile binding site by which it may become attached to various types of surfaces including cell membranes (Müller-Eberhard et al, 1966). The labile binding site is thought to arise

by the cleavage of an internal thiolester bond to release a free acyl group. This could then form a covalent ester bond by interacting with a hydroxyl group, possibly in a sugar residue, on the activating surface (Law and Levine, 1977; Capel et al, 1978; Law et al, 1979a, 1979b; Sim et al, in press). In the case of IgG aggregates, it is thought that the covalent bond is an amide linkage involving an amino group in the Fd fragment of IgG (Gadd and Reid, 1981b).

The labile binding site of C3b is located in the C3d fragment of C3 (Law et al, 1979a), the fragment which remains bound to the cell surface after cleavage and inactivation of C3b by its control protein, C3b INA. The role of C3b in the complement pathway is multifarious: it is able to bind (i) factor B to form the alternative pathway C3 convertase, (ii) C5 to allow its cleavage by a C5 convertase, (iii) C5b6 in order to concentrate the terminal components on the cell membrane and permit their more efficient utilisation, (iv) β1H which competitively inhibits the binding of other components and enhances the proteolytic activity of C3b INA.

Because of its extremely important central role in both the complement pathway and other biological functions, the activity of C3 is finely controlled to prevent C3 depletion. This control is firstly inherent in the short half-life of the labile binding site in nascent C3b. The second form of control is through the control proteins C3b INA and β1H. C3b INA is an enzyme of 88,000 mol. wt. and splits the α' chain of C3b and so destroys its biological

activity (Tamura and Nelson, 1967; Gitlin et al, 1975; Nagasawa and Stroud, 1977). The importance of this inhibitor is exemplified in the C3 and factor B depletion, and extreme susceptibility to bacterial infections of an individual lacking C3b INA (Alper and Rosen, 1971). β1H is a glycoprotein of 150,000 mol. wt. and acts as an accelerator of C3b INA (Whaley and Ruddy, 1976; Weiler et al, 1976).

1.2.1.4 Activation of the alternative pathway

The initiation of the alternative pathway is mediated by the binding of C3b via its labile binding site to a surface which allows it to be protected from the regulatory action of the control proteins, C3b INA and β1H (Fearon and Austen, 1977a; 1977b). Low level turnover of C3 occurs in the unstimulated organism and this is probably the source of the initial C3b (Nicol and Lachmann, 1973; Lachmann and Halbwachs, 1975). Once stabilised, C3b can bind factor B, and following activation by factor $\bar{D}$, the complex forms the alternative pathway C3 convertase.

A wide variety of substances are able to activate the alternative pathway, and the extensive literature which has accumulated has been reviewed by Götze and Müller-Eberhard (1976), and Reid and Porter (1981). The activating substances can be broadly divided into two groups - immunoglobulin and non-immunoglobulin.

(1) Immunoglobulin activators

Human IgA, IgD and IgE and some rabbit, guinea-pig, ruminant and human IgG's can activate the alternative pathway when aggregated in immune complexes or by chemical procedures. C3b binds covalently to the $F(ab')_2$ fragments of immunoglobulin (Takahashi et al, 1976; Gadd and Reid, 1931a). The integrity of the inter-heavy chain disulphide bond is essential for alternative pathway activation (Gadd and Reid, 1981a) and could be critical in maintaining a conformation which allows C3b binding and protection.

Specific antibody can also enhance the rate of alternative pathway activation by non-immunoglobulin activators (Polhill et al, 1978; Sissons et al, 1979; Nelson and Ruddy, 1979; Edwards et al, 1980). For example, it has been observed that human IgG will enhance the lysis of rabbit erythrocytes by the alternative pathway (Nelson and Ruddy, 1979), and specific antibody to type III, group B streptococci will enhance their phagocytosis by the alternative pathway (Edwards et al, 1980). In this latter case, it is possible that the role of antibody is to mask sialic acid residues on the surface of the streptococci which otherwise are poor activators of the alternative pathway. Enzymatic removal of the sialic acid produced the same enhancing effect.

(2) Non-immunoglobulin activators

The alternative pathway can be activated by a variety of polysaccharides (zymosan, inulin, endotoxin from gram-negative bacteria), polyanions (dextran sulphate, DNP-

albumin, polyvinyl sulphonate), other constituents of microbial cell walls (teichoic acid, peptidoglycan, lipopolysaccharides), intact micro-organisms (E. coli, certain parasites) and mammalian cells (rabbit erythrocytes, neuraminidase-treated sheep erythrocytes, measles virus infected HeLa cells, peripheral B lymphocytes and Raji cells). It is thought that the essential common feature which links these diverse activators is the presence of some form of repeating structure (Burger et al, 1975, 1978; Capel et al, 1978). Certain cell surface components have also been implicated in modulating activation - sialic acid (Fearon, 1978; Edwards et al, 1980) and sulphated glycosaminoglycans (Kazatchkine et al, 1979) in nonactivators, and lipopolysaccharides (Pangburn et al, 1980), particularly those carrying a positive charge (Cunningham et al, 1979) in activators. Thus it seems likely that several structural features may be involved.

The interaction of the labile binding site on C3b with the activating surface is thought to consist of a primary interaction via a covalent bond (section 1.2.1.3) and secondary "discriminatory" interactions involving other surface structures (Schreiber et al, 1978).

1.2.1.5 Alternative pathway C3/C5 convertases

Once C3b has bound to the surface of an activator, it can bind factor B in a 1:1 stoichiometry in the presence of Mg^{2+} ions (Müller-Eberhard and Götze, 1972; Vogt et al, 1977). It has been reported that the convertase in this form has limited C3 cleaving ability

(Fearon and Austen, 1975a; Day et al, 1976; Daha et al, 1976; Vogt et al, 1977; Brade et al, 1977a) but full activity is not expressed until factor B is cleaved by factor $\bar{D}$ (Götze and Müller-Eberhard, 1971; Daha et al, 1976; Lesavre et al, 1979; Kerr, 1979). Once activated, the convertase can cleave further C2 to C3b which can then be deposited on the activator surface and form more C3bB. This amplification phase is the C3 feedback cycle of the alternative pathway and is finely controlled by C3b INA and β1H so that C3 exhaustion does not occur (Alper et al, 1972; Whaley and Thompson, 1978; Fearon and Austen, 1977a, 1977b).

The half-life of $\overline{C3b,B}$ is 15-20 min at 37°C (Fearon et al, 1973) but the complex can be stabilised by properdin (Schreiber et al, 1978), a glycoprotein of mol. wt. 220,000, or by C3NeF, a conformational antibody to the C3 convertase (Davis et al, 1977; Daha etaal, 1978). Once Bb decays from C3b the convertase becomes inactive, but the C3b can be reused to generate $\overline{C3bB}$ in the presence of factor $\bar{D}$, Mg^{2+} ions and fresh factor B.

The C3b generated by the convertase can also combine with already-formed $\overline{C3bB}$, after which the specificity of the convertase is altered so that it will cleave the α-chain of C5 (Cooper and Müller-Eberhard, 1970). The mechanism by which this change of specificity occurs is not clear, but it is probable that the binding of nascent C3b in or near the vicinity of the convertase allows reversible interaction of C5 with C3b, and that C5

cleavage occurs only after this interaction has occurred (Vogt et al, 1978; 1979; von Zabern et al, 1980; Strunk and Giclas, 1980). If such a mechanism existed, this would suggest that the C3 and C5 convertases have the same structure (i.e. $C\overline{3bB}$) and that the two substrates were C3 and C3b-bound C5 respectively (Vogt et al, 1978). It appears that more than one additional C3b molecules are necessary to form the C5 convertase, and that these additional molecules may be more susceptible to destruction by C3b INA (Reid and Porter, 1981).

1.2.2 The classical pathway

The classical pathway (Fig. 1.3) consists of eleven major proteins, C1q, C1r, C1s and C2-C9, together with the control proteins $C\bar{1}$ INH, C4 binding protein, β1H and C3b INA. Although important analogies can be drawn between the activation of the classical and alternative pathways (Fig. 1.1) major differences occur, both in the activation of $C\bar{1}s$ (functionally analogous to factor $\bar{D}$ in the alternative pathway), and in the absence of a C3b feedback system.

1.2.2.1 C1q

C1q (mol. wt. 400,000) is a glycoprotein and appears under the electron microscope to be composed of six globular regions attached by connecting strands to a fibril-like central portion (Shelton et al, 1972; Knobel et al, 1975; Brodsky-Doyle et al, 1976). It is composed of 18 polypeptide chains, 6A, 6B and 6C, each of mol. wt. 23,000-24,000 (Reid et al, 1972; Reid and Porter, 1976).

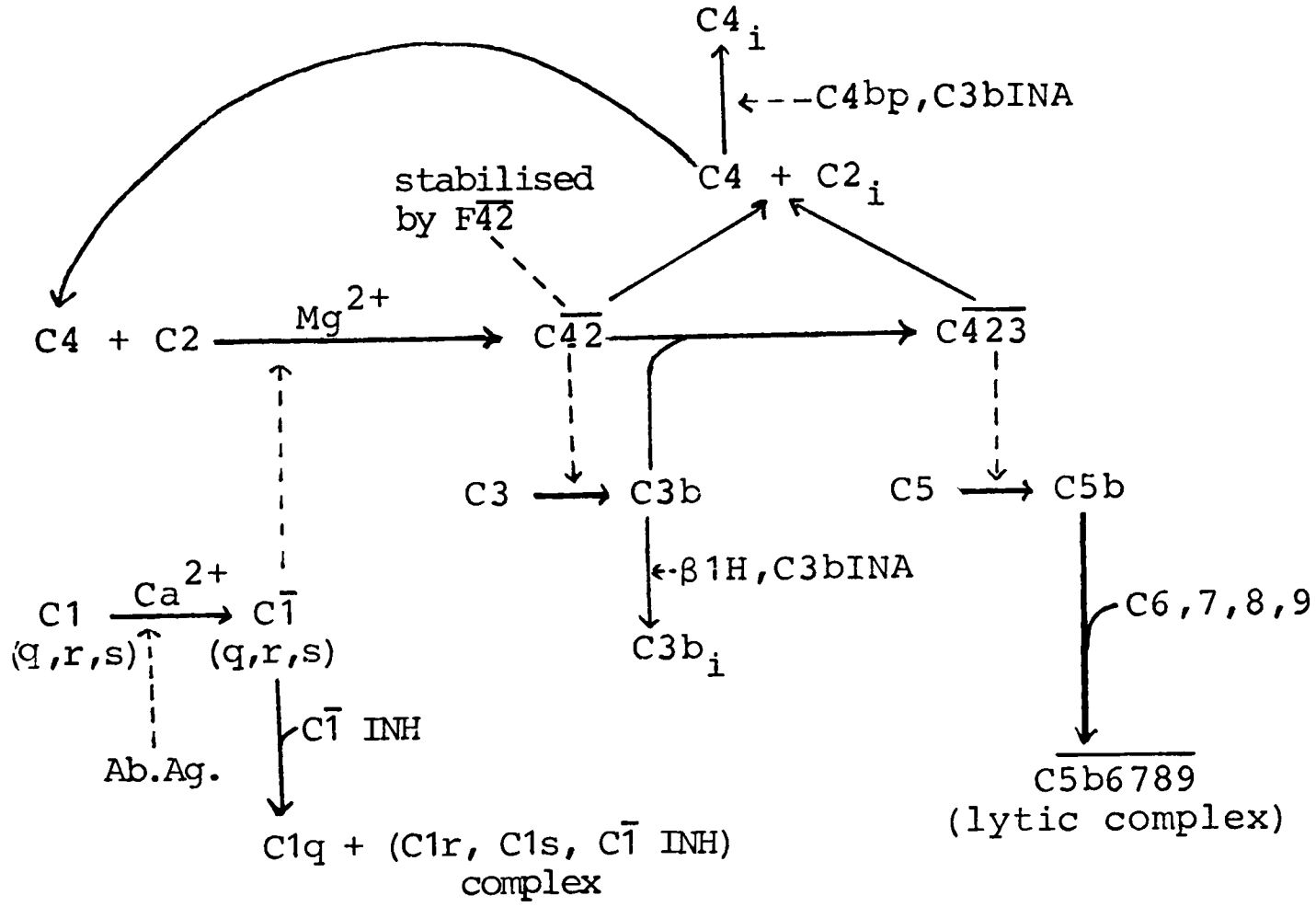


Fig. 1.3 The classical pathway

A unique feature of the molecule is that 40% of its structure is collagen-like (Reid, 1976) and contains the sequence $-(gly-X-Y)_n-$, where X is often proline and Y is often hydroxyproline or hydroxylysine. This type of sequence forms a stable triple helix and is responsible for the fibril-like structure seen in the electron microscope. A break in the triplet sequence is thought to correspond to the points at which the central fibril divides to form the six connecting strands to the globular regions (Porter and Reid, 1979). Following digestion with pepsin, the globular regions are lost but the connecting strands and central fibril remain intact (Reid, 1976). Conversely, the collagen-like regions are partially digested by collagenase, leaving the globular heads intact (Knobel et al, 1974; Hughes-Jones and Gardner, 1979; Pâques et al, 1979). The proposed structure of C1q (Fig. 1.4) has been reviewed by Porter and Reid (1978, 1979).

C1q interacts via its globular regions with the Fc portion of immunoglobulin (Knobel et al, 1974; Wautier et al, 1977). It is thought that the binding site for immunoglobulin may consist of a linear sequence of amino acids (Isenman et al, 1975) but this has not been clearly demonstrated.

1.2.2.2 C1r and C1s

C1r and C1s are both glycoproteins of mol. wt. 85,000. They are very similar in gross structure and on activation are cleaved to give two polypeptide chains

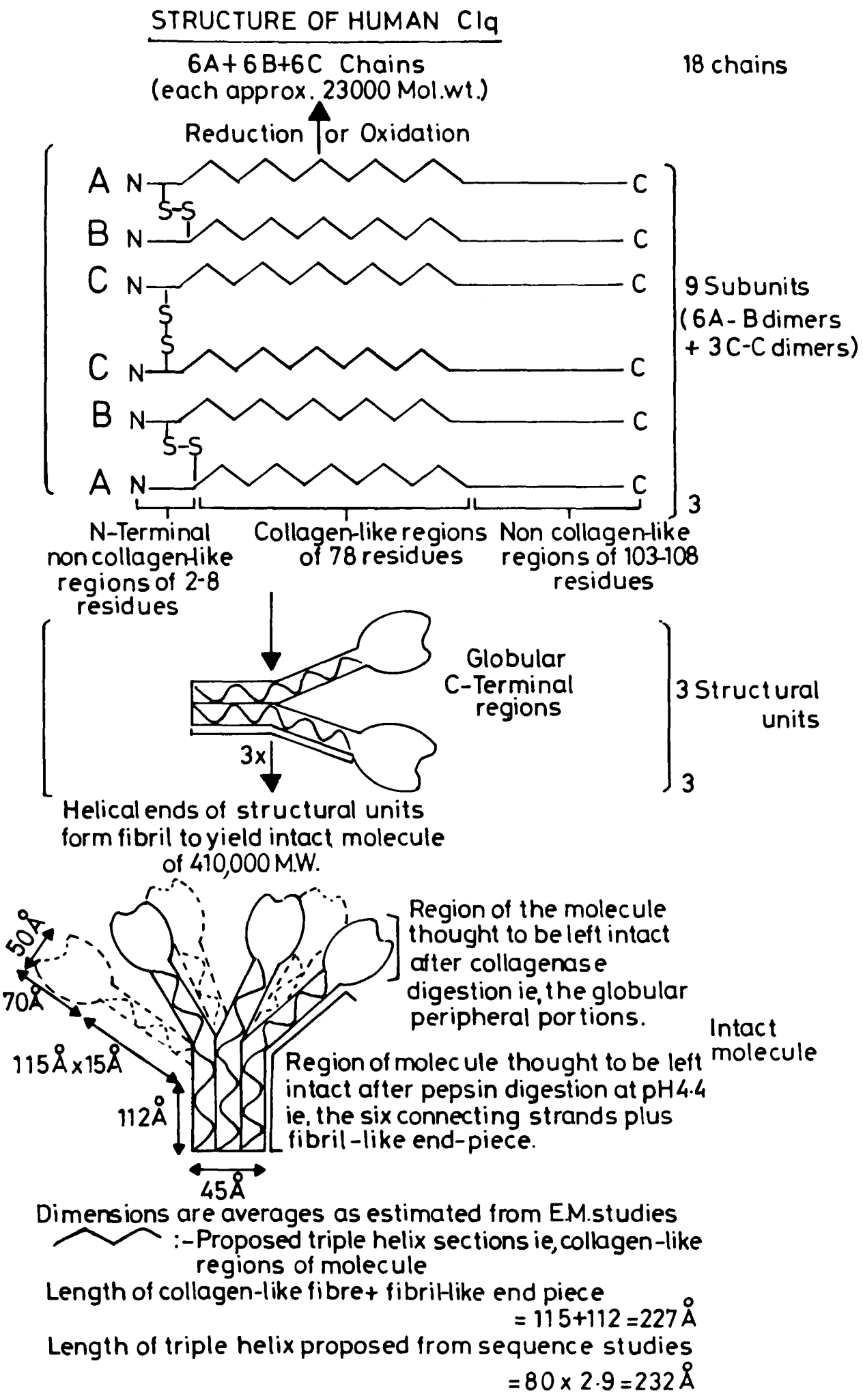


Fig. 1.4 Model for the structure of C1q

(from Porter and Reid, 1978)

of about 27,000 (b) and 56,000 (a) mol. wt. linked by disulphide bonds. In both cases the smaller b chain has serine protease activity which is inhibitable by iPr_2P-F. The N-terminal sequence of the b chain in both cases (Fig. 6.1) shows strong homology to those of other serine proteases (Sim et al, 1977).

Although structurally similar, $C\bar{1}r$ and $C\bar{1}s$ differ markedly in their enzyme specificity. C1s is the only known substrate for $C\bar{1}r$, whereas $C\bar{1}s$ will cleave both C2, C4 and several synthetic lysyl and tyrosyl esters (Sim et al, 1977).

In nondissociating conditions in the presence or absence of EDTA, purified C1r forms a dimer whereas $C\bar{1}s$ is monomeric (Ziccardi and Cooper, 1976; Sim et al, 1977). In the presence of Ca^{2+} ions, a tight tetrameric complex containing 2C1r and 2C1s molecules forms. Recent evidence suggests that the unactivated tetramer binds 4 Ca^{2+} ions and the activated tetramer, 5 Ca^{2+} ions (Villiers et al, 1980). In $C\bar{1}r$ at least, the Ca^{2+} binding site appears to be in the noncatalytic a chain (Villiers et al, 1980).

The C1r-C1s complex binds to the collagenous regions of C1q (Reid et al, 1977). The complex of $C\bar{1}$(q,r,s), which is stabilised by Ca^{2+} ions, can be dissociated by $C\bar{1}$ INH, a stoichiometric inhibitor of $C\bar{1}r$ and $C\bar{1}s$. $C\bar{1}$ INH is a single chain glycoprotein of mol. wt. 100,000 and has an unusually high (40%) carbohydrate content (Haupt et al, 1970; Harpel et al, 1975). It binds to the catalytic chains of

$C\bar{1}r$ and $C\bar{1}s$ and causes their dissociation from C1q (Harpel et al, 1975; Sim et al, 1979; Ziccardi and Cooper, 1979).

1.2.2.3 C2

C2 has a mol. wt. of about 100,000. On activation by $C\bar{1}s$, it is split into two fragments, C2a (mol. wt. 73,000) which forms part of the classical pathway C3 convertase, and C2b (34,000 mol. wt.). There have been conflicting reports of C2 inhibition by iPr_2P-F (Polley and Müller-Eberhard, 1968; Cooper, 1975; Medicus et al, 1976; Kerr, 1980) but it appears to be a novel serine protease probably similar to factor B, with the active site in the larger C2a fragment. N-terminal sequence data on C2a indicates a lack of homology with typical serine proteases (Kerr, 1979).

1.2.2.4 C4

C4 is a glycoprotein of mol. wt. 200,000. It is synthesized as a single peptide chain (Hall and Colten, 1977) which is hydrolysed to give a three chain molecule held together by disulphide bonds and strong noncovalent interactions. The three chains, α, β and γ have mol. wts. of 93,000, 78,000 and 30,000 respectively and appear to be aligned in the precursor molecule in the order, β-α-γ (Abraham et al, 1980; Goldberger and Colten, 1980; Parker et al, 1980).

Activation of C4 is brought about by $C\bar{1}s$ and results in the cleavage of a peptide (C4a) derived from the N-terminal of the α-chain. C4a has anaphylatoxic properties

(Gorski et al, 1979b) and shows some structural similarity to C3a and C5a (Gorski et al, 1979a, 1979b; Booth et al, 1979). The larger fragment, C4b, contains a labile binding site which enables it to bind covalently to cell membranes and IgG (Müller-Eberhard and Lepow, 1965). The half-life of the labile binding site is short, and even under optimal conditions, only 5-10% of the available C4 is bound to cells coated with antibody and C1 (Cooper and Müller-Eberhard, 1968).

C4b is inactivated by C3b INA in association with C4 binding protein (C4 bp) which plays an analogous role to β1H in C3b regulation (Scharfstein et al, 1978; Fujita and Nussenzweig, 1979). It has a mol. wt. of 540-590,000 and is composed of about 8 polypeptide chains.

1.2.2.5 Activation of the classical pathway

The classical pathway is activated by aggregated IgG or IgM. Neither monomeric IgG nor aggregates containing IgA, IgD or IgE are effective in causing C1 activation (Augener et al, 1971; Spiegelberg, 1974). Certain antibody independent pathways of C1 activation are also known (Hughes-Jones et al, 1977; Reid, 1980).

When aggregated immunoglobulin is the activator, the globular regions of C1q interact with the Fc region of the antibody (Reid et al, 1972; Knobel et al, 1974), and one molecule of C1q will bind 12-18 molecules of IgG (Schumaker et al, 1976). In IgG it is the C_H2 domain that is involved (Kehoe and Fougereau, 1969; Kehoe et al, 1974;

Colomb and Porter, 1975; Yasmeen et al, 1976), and in IgM, the C_H4 domain (Hurst et al, 1975). The possible mechanisms of binding to C1 have been reviewed by Metzger (1974, 1978). An associative model is favoured in which antibody aggregation results in the presentation of multiple Fc regions and consequently, stronger interaction with C1q.

The mechanism of C1q activation is also unclear. In solution, C1q, C1r and C1s are only weakly associated, but once C1q has bound to aggregated IgG, the C1 complex becomes tightly bound and only dissociable in EDTA. Thus Reid et al (1977) have suggested that binding to antibody causes a conformational change in C1q which leads both to tighter binding to the C1r-C1s tetramer and to activation of C1r. Activation of C1r will only occur in the assembled C1 complex. It has been proposed that following the putative conformational change in C1q, an active site is exposed in the unsplit C1r molecule. This would then cleave itself or an adjacent C1r molecule to yield $C\bar{1}r$ which would then activate C1s (Dodds et al, 1978; Lin and Fletcher, 1980).

1.2.2.6 Classical pathway C3/C5 convertases

Once $C\bar{1}s$ is activated, it will cleave C4 and C2 in solution. C4b, like C3b, has a labile binding site by which it can bind to cell surfaces or to aggregated IgG (Müller-Eberhard and Lepow, 1965). It is the C4 which interacts with antibody which is involved in C3 convertase formation (Goers and Porter, 1978). The interaction is via

a covalent ester bond between the C4d fragment of the α' chain of C4b (R.D. Campbell, unpublished results) and the Fd fragment of IgG (Campbell et al, 1980). It is possible that as in C3, an acyl group is released when C4 is activated, and that this group reacts with an amino acid in the immunoglobulin (Campbell et al, 1980).

Once C4b has bound, it is thought that whole C2 associates with C4b in the presence of Mg^{2+} ions, and is then cleaved in the complex by $C\bar{1}s$ (Dodds and Porter, 1979). $C\bar{1}s$ does not appear to form a part of the convertase.

$C\overline{42}$ is a labile complex with a half-life of 7 min at 37°C (Lachmann, 1973). On decay of the convertase C2a dissociates, but the C4b can be reused to form fresh convertase on addition of uncleaved C2 and $C\bar{1}s$ (Müller-Eberhard et al, 1967). $C\overline{42}$ can be stabilised by $F\overline{42}$, a $C\overline{42}$ "nephritic factor" (Daha et al, 1980) or by prior treatment of C2 with iodine ($C2^{oxy}$) which is thought to modify C2 by oxidation of critical sulphydryl groups (Polley and Müller-Eberhard, 1967).

$C\overline{42}$ will cleave C3 to C3b which can then activate the alternative pathway feedback mechanism (section 1.2.1.5) or take part in C5 conversion. It is thought that the C3 and C5 convertases are identical, and that the substrate is C3b-bound C5 (Vogt et al, 1978; 1979; Isenman et al, 1980). This situation is analogous to that proposed for the alternative pathway (section 1.2.1.5) except that probably only one C3b molecule is required.

1.2.3 The terminal components (C5-9)

The terminal stages of immune haemolysis have been reviewed recently by Mayer et al (1979) and Boyle and Borsos (1980). The structural data for components C5-9 have been reviewed by Fothergill and Anderson (1978); Porter (1979) and Reid (1980).

Following activation of C5, the membrane attack complex (MAC) appears to form by self-assembly without the involvement of proteases. C5 activation occurs by its cleavage by a C5 convertase to form C5a and C5b. Nascent C5b has a hydrophobic binding site by which it can attach to an acceptor (Müller-Eberhard, 1975). The activated state of C5b is stabilised by C6 (Goldlust et al, 1974) and once C5b6 combines with C7, the complex formed can penetrate the membrane lipids (Hammer et al, 1976; Podack and Müller-Eberhard, 1978; Podack et al, 1979; Podack et al, 1980). On addition of C8, the complex becomes further embedded in the hydrophobic layers of the membrane, and marked changes in ion permeability and the organisation of the lipid bilayer occurs (Michaels et al, 1976; Giavedoni and Dalmasso, 1976; Esser et al, 1979; Podack et al, 1980), and some membrane damage is possible (Stolfi, 1968; Lachmann et al, 1970). C9 (6 molecules per 2C5b-8) is required to induce formation of the stable dimeric MAC and to cause complete cell lysis. The typical membrane lesion consists of a raised, electron-lucent, ring-like structure surrounding an electron dense centre. It is thought that the MAC either forms a cylinder

of complement proteins around a central pore (Mayer, 1972; Dourmashkin, 1978; Bhakdi and Tranum-Jensen, 1978) or that the MAC causes reorientation of the lipid molecules so that a transmembrane channel is formed around the complex, rather than through its centre (Podack and Müller-Eberhard, 1978). The net result is breakdown of membrane semi-permeability and cell lysis.

1.3 SERINE PROTEASES

Early information on the structure and function of serine proteases was obtained largely from bovine chymotrypsinogen (CTG) and chymotrypsin (CT) and has been reviewed by Kraut (1971); Blow (1971); Hess (1971); and Bender and Killheffer (1973). However, several families occur within the serine proteases, and de Haën *et al* (1975) have recognised four such families among vertebrates : trypsin, chymotrypsin, elastase and thrombin, named after their best-known members.

Bovine CTG-A is probably the best studied serine protease zymogen and the numbering and amino acid residues mentioned in the following sections are for this enzyme unless stated otherwise. The structure of bovine CT has also been used here as a basis from which to discuss other vertebrate serine proteases.

1.3.1 Structure/function relationships in serine proteases

Tables of the alignment of amino acid sequences of serine proteases have been published by de Haën *et al* (1975) and Dayhoff (1979). A comparison of the primary structures shows that the amino acid sequence is highly conserved in the regions of the N-terminal, the catalytic site and other residues important in either substrate binding or in maintaining the active configuration. Comparisons of sequence data for the enzymes CT, trypsin, elastase and thrombin show that 45% of their "internal" residues are invariant, compared with only 6% of the "external" residues (Hartley and Shotton, 1971).

The higher structure of several serine proteases including enzyme-inhibitor complexes, has been investigated by X-ray-diffraction (usually at 2.5-2.7 Å resolution), optical rotatory dispersion and circular dichroism. These data are reviewed in "The Enzymes" vol III (ed. P.D. Boyer, Academic Press, 1971) and by Bender and Killheffer (1973); Stroud et al (1975); and Kraut (1977).

1.3.1.1 Gross topography

CT is a slightly ellipsoid molecule which is flattened a little at the active site. The active site consists of a side chain binding pocket located beneath the catalytic site which contains the reactive serine residue (Steitz et al, 1969). CT contains little regular secondary structure. Two short regions of α-helix are found at residues 235-245 and 164-173, and antiparallel β-sheet structure is common but not regular. Residues 27-112 and 133-230 form two series of six antiparallel β-sheets which are hydrogen-bonded in such a way as to form two distorted cylinders containing cores of hydrophobic residues (Birktoff et al, 1970). The two cylinders of β-structure lie over each other, oriented approximately at right angles. The gross structures of trypsin (Stroud et al, 1974) and elastase (Watson et al, 1970; Shotton and Watson, 1970) are remarkably similar to that of CT, although the identity of amino acid sequence among the three proteins is only 24% (Stroud et al, 1974). This similarity of gross structure is achieved by compensating amino acid substitutions in spatially adjacent regions of the molecule so that the tertiary structure is relatively unaffected (Watson et al

1970). However, small differences do occur. X-ray crystallographic data on thrombin are not available, but predicted structures of the catalytic B chain suggest that it will probably be similar to those of other major families of vertebrate serine proteases (Robson, 1980).

Vertebrate serine proteases contain at least three conserved disulphide bonds (Fig. 1.5). These are cys 42-58 (the "his" loop), cys 168-182 (the "met" loop) and cys 191-220 (the "ser" loop). Additional disulphide bonds may also be present.

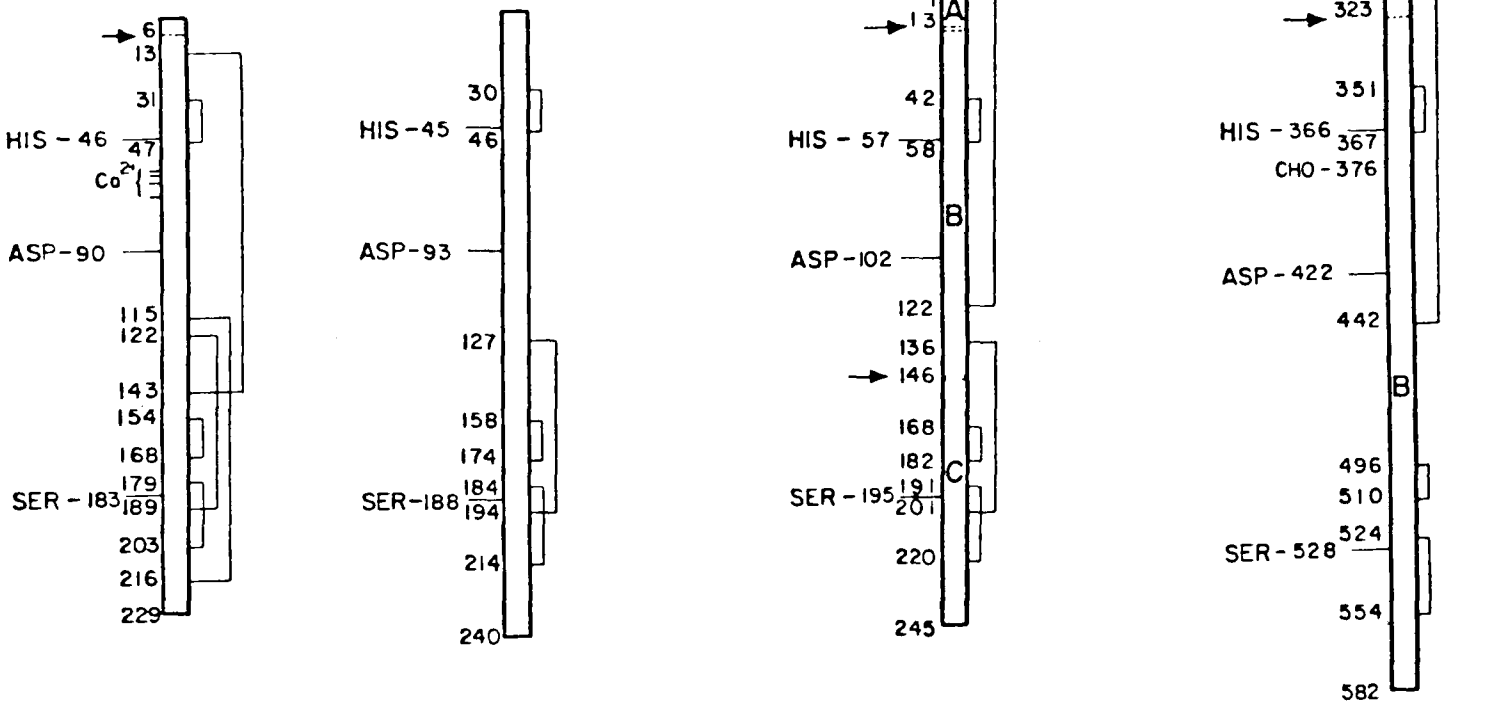
1.3.1.2 Structure of the active site

The structure of the active sites of serine proteases has been studied using the native enzyme, enzyme-inhibitor complexes, enzyme-virtual substrate complexes and acyl enzyme derivatives (reviewed by Blow, 1971; Stroud *et al*, 1975; Kraut, 1977). Kraut (1977) has defined five features common to the active sites of serine proteases:

(i) the extended polypeptide binding sites on the acyl group side of the susceptible peptide bond,

(ii) sites for binding, with varying degrees of specificity, the side chains of a polypeptide substrate,

(iii) a site for binding the substrate on the side of its leaving group,



Trypsinogen
Bovine
Figure 20

Elastase
Pig
Figure 21

Chymotrypsinogen A
Bovine
Figure 22

Prothrombin
Bovine
Figure 24

Fig. 1.5 Molecular architecture of some vertebrate serine proteases (from Dayhoff, 1979)

Figures are drawn approximately to scale. The active site residues (his, asp, ser) and the disulphide-bond patterns are shown. Arrows indicate sites/cleaved during activation; letters (A, B, C) indicate chain structure of activated enzyme; CHO = carbohydrate binding residues; Ca^{2+} = sites which bind calcium (represented by stippled region in prothrombin); shaded areas in prothrombin are kringle regions which, together with the calcium binding site, are lost on activation.

(iv) a site (the "oxyanion hole") for binding the carbonyl oxygen atom of the susceptible peptide bond when the carbonyl group is in a tetrahedral configuration,

(v) the charge relay system including the reactive serine side chain, which forms a covalent bond with the carbonyl carbon of the susceptible peptide bond.

The residues involved in the specificity sites of CT, trypsin and elastase are shown diagramatically in Fig. 1.6.

The amino acid residues extending from the cleavage site toward the N-terminal of the substrate (the acyl group side) are named P_1, P_2, P_3..... and those extending towards the C-terminal (the leaving group side), P'_1, P'_2, P'_3..... The corresponding binding sites on the enzyme are denoted S_1, S_2, S_3..... and S'_1, S'_2, S'_3..... respectively.

(1) Binding on the acyl group side of the substrate

Serine proteases contain a conserved sequence, ser_{214} - trp_{215} - gly_{216}, which forms a typical hydrogen-bonded antiparallel β-structure with the polypeptide chain of a suitable substrate or inhibitor (Segal et al, 1971). Whereas, P_1 and P_3 of the substrate interact with ser_{214} (S_1) and gly_{216} (S_3) respectively, P_2 interacts with residue 99, which in CT is isoleucine. This residue is not conserved in serine proteases but does tend to be hydrophobic (Kraut, 1971).

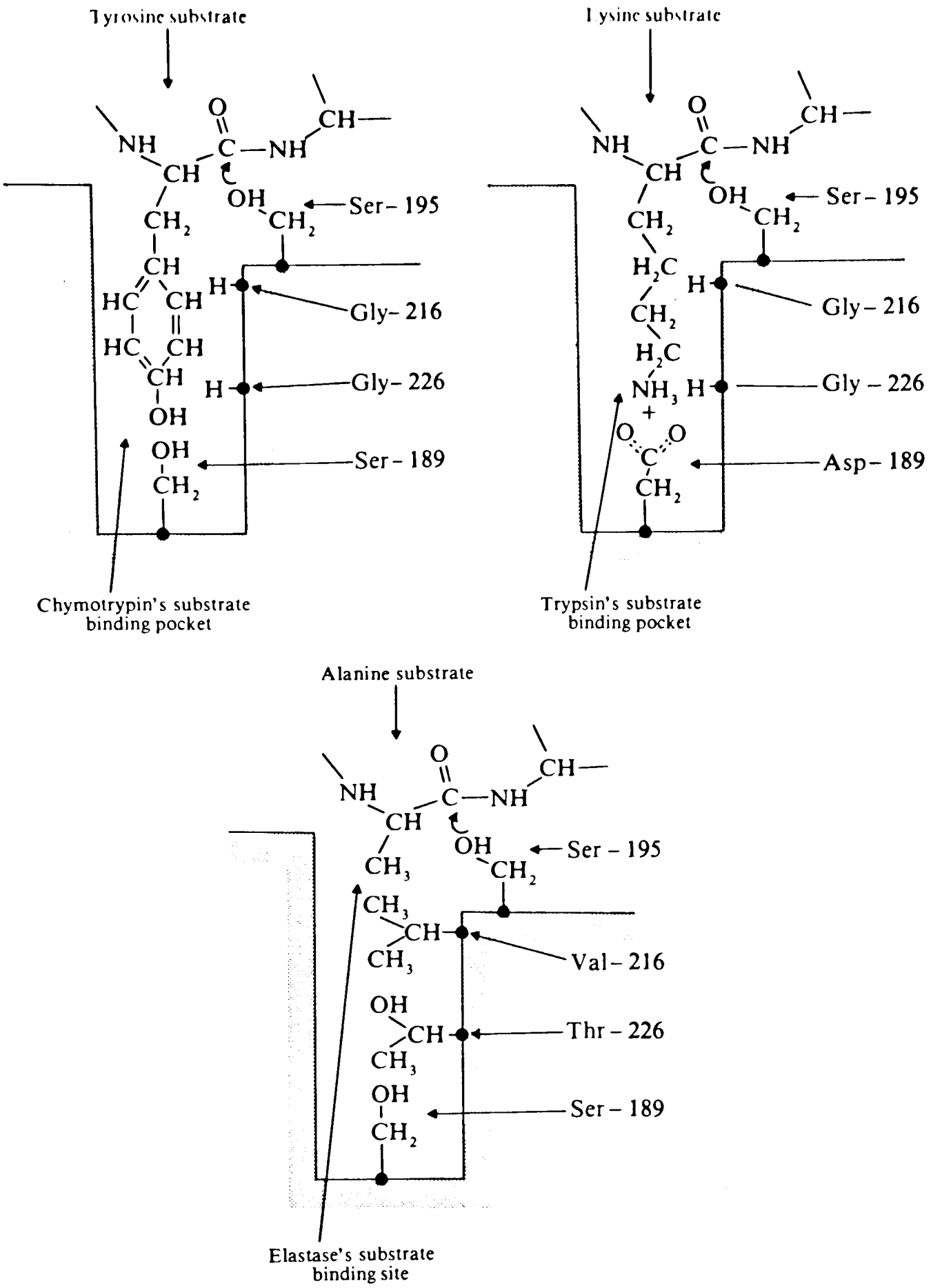


Fig. 1.6 Specificity sites of chymotrypsin, trypsin and elastase (from Blake, 1979)

(2) The specificity site

Besides the hydrogen-bonding of the polypeptide backbone of the substrate, the amino acid side chains are also bound with varying degrees of specificity. CT is most effective in the hydrolysis of peptide bonds following an aromatic sidechain at P_1; trypsin and thrombin are specific for hydrolysis following arginine or lysine; and elastase cleaves small, uncharged, non-aromatic side chains, especially alanine. This specificity for P_1 is explained by the molecular structure of the specificity pocket which lies below the reactive serine residue (Fig. 1.6) and in which P_1 is bound (Steitz, 1969).

In trypsin and CT, the chain segment of residues 184-193 and 217-223 form the specificity pocket, together with the disulphide bond cys 191-220. One wall of the pocket is made up of residues 214-216 (S_1 - S_3) and the other wall, of the peptide links from residues 190-192. Residue 192 (met in CT, gln in trypsin) acts as a flexible lid on the pocket and aids substrate binding (Steitz, 1969; Krieger et al, 1974). Residue 189 is at the base of the pocket. The result of this arrangement is that in CT, a planar aromatic P_1 side chain can be bound, whereas in trypsin and thrombin the substitution of asp for ser_{189} allows a lys or arg side chain to be stabilised. In elastase, the base of the pocket is occluded both by the substitution of val at residue 216 which blocks the entrance to the pocket, and of thr for gly_{226}. Thus, only small side chains can be bound at S_1 in elastase and, in contrast to the situation with trypsin and CT, P_2 and P_3

markedly enhance the reactivity of elastase toward the substrate.

(3) Binding on the leaving group side of the substrate

The interactions on the leaving group side are less well established. They do not appear to be as specific and are possibly weaker than on the acyl group side (Kraut, 1977). Preferences for specific side chains are not well established and antiparallel β-structure between the S' and P' regions may or may not occur.

(4) The oxyanion binding site

When P_1 is bound in the specificity pocket, its carbonyl carbon forms a tetrahedral intermediate (see Fig. 1.7a, section 1.3.2) attached to four ligands, P_1, P'_1, the carbonyl oxygen and the hydroxyl oxygen (O_γ) of the reactive serine in the enzyme. In this situation, the carbonyl oxygen forms hydrogen-bonds with the amide hydrogens of ser_{195} and gly_{193} (Henderson, 1970).

(5) The reactive serine and the charge relay system

The characteristic feature of serine proteases is the presence in the active site of a serine residue with unusually strong nucleophilic properties. Indeed, the diagnostic test for a serine protease has been the inhibition of its enzymic activity by the inhibitor, iPr_2P-F, with which the active site serine forms a covalent complex (Kraut, 1977). The source of the extreme nucleophilicity of this serine residue was long thought to be the charge relay system (Blow, 1969) in which the

residues asp_{102}, his_{57} and ser_{195} were hydrogen-bonded in that order. Although the charge-relay system certainly exists and is conserved among all the serine proteases studied to date, its importance in contributing to the nucleophilicity of the active site serine has been challenged (Kraut, 1977; Matthews et al, 1977). In the model of Matthews et al (1977), asp_{102} which is buried in a relatively hydrophobic environment, serves to correctly orient his_{57} with respect to accepting a proton from ser_{195}. The apparent nucleophilicity of the serine residue is attributed to the inherent instability of the tetrahedrally-distorted carbonyl bond being cleaved. However, the role of his_{57} as a proton acceptor from ser_{195} is not disputed, and the problem can be viewed in terms of the timing of proton removal, either before, during or following attack of ser_{195} on the substrate (Lawson, 1980).

1.3.1.3 Activation of the zymogen

The conversion from zymogen to active enzyme is achieved by the cleavage of at least one peptide bond to release a new N-terminal amino acid which is usually isoleucine. To maintain an active configuration in CT or trypsin, the N-terminal amino group must be protonated and form an ion pair with the carboxylate group of asp_{194} in the active site (Oppenheimer et al, 1966; Kraut, 1971; Blow, 1971). This is achieved by the new N-terminal ile_{16} moving 11.3 $\mathring{A}$ by rotating through 180^{o} and becoming buried just below the surface of the molecule in an ion pair with asp_{194}. This movement results in further small

conformational changes which include extension of the polypeptide backbone in the active site region (section 1.3.2.2).

The exact reasons for the lack of activity in the zymogen have not been satisfactorily explained, and may vary in different proteases. The gross structures of zymogen and activated protease are similar (Bender and Killheffer, 1973; Stroud et al, 1975; Kraut, 1977). The charge relay system is almost intact in the zymogen, except for a slight twist in the side chain of his_{57} (Kraut, 1971). Whereas this slight distortion may be sufficient to prevent enzyme activity, it is also pertinent that stabilisation of the tetrahedral intermediate by hydrogen-bonding to gly_{193} cannot occur in the zymogen. In CTG and trypsinogen, the active site is essentially preformed and is not sterically obstructed (Freer et al, 1970; Kraut, 1971; Wright, 1973). The zymogens show weak intrinsic enzymic activity (Dlouhá and Keil, 1969; Kay and Kassel, 1971; Gertler et al, 1974; Kerr et al, 1975) and they appear to interconvert between "active" and "inactive" configurations (Bode and Huber, 1976). However, substrate binding is impaired and the specificity pocket is distorted (Freer et al, 1970; Kassel and Kay, 1973; Gertler et al, 1974). In proelastase, the substrate may be bound but cleavage is not possible (Hartley and Shotton, 1971). In the active enzyme, protonation of its N-terminal valine does not appear to be obligatory for enzyme activity, although it is necessary for this residue to be buried (Hartley and Shotton, 1971). Studies on thrombin and factor X also suggest that alternative mechanisms for stabilising

asp_{194} may exist (de Haën et al, 1975).

The stabilisation of an active conformation is of particular interest when considering the activity of the complement proteases, since the zymogens of some of these also appear to adopt catalytically active forms.

1.3.2 Enzymic activity

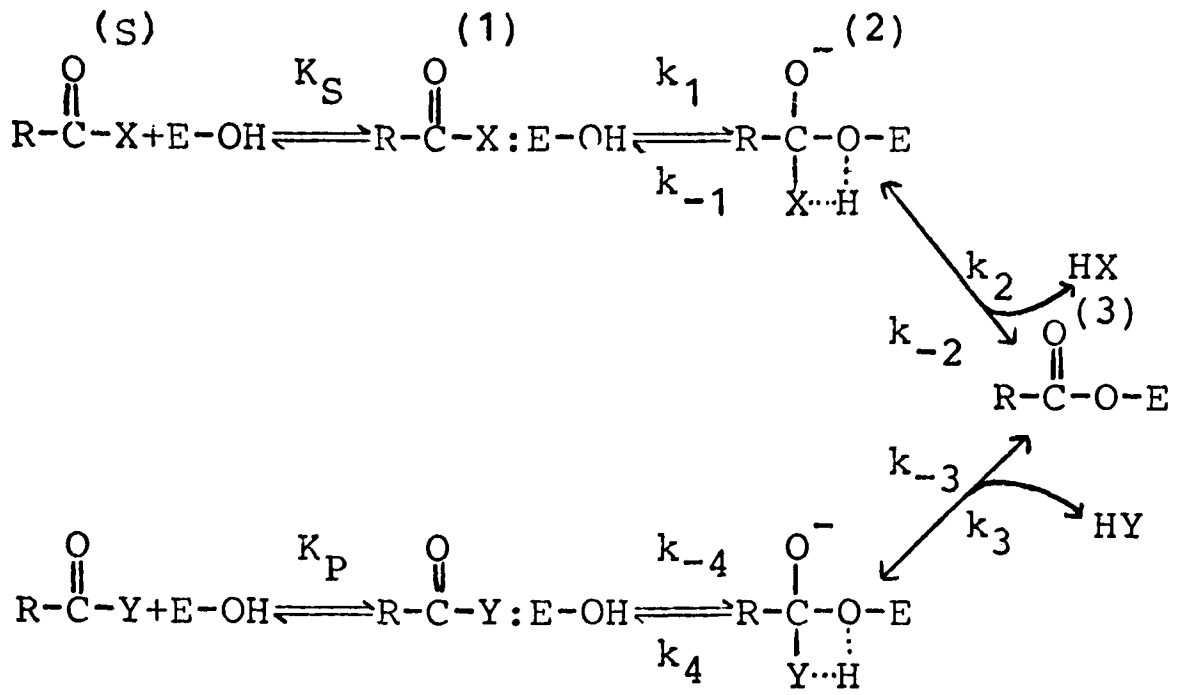
The catalytic mechanism of peptide bond cleavage by serine proteases has been elucidated by both structural and kinetic studies on the native and inhibited enzyme (Hess, 1971; Bender and Killheffer, 1973; Stroud et al, 1975; Kraut, 1977).

The general mechanism of catalysis (Fig. 1.7a) may be represented by a reaction sequence in which product release is regarded as a reversal of substrate binding (Kraut, 1977).

1.3.2.1 Groups involved in catalysis

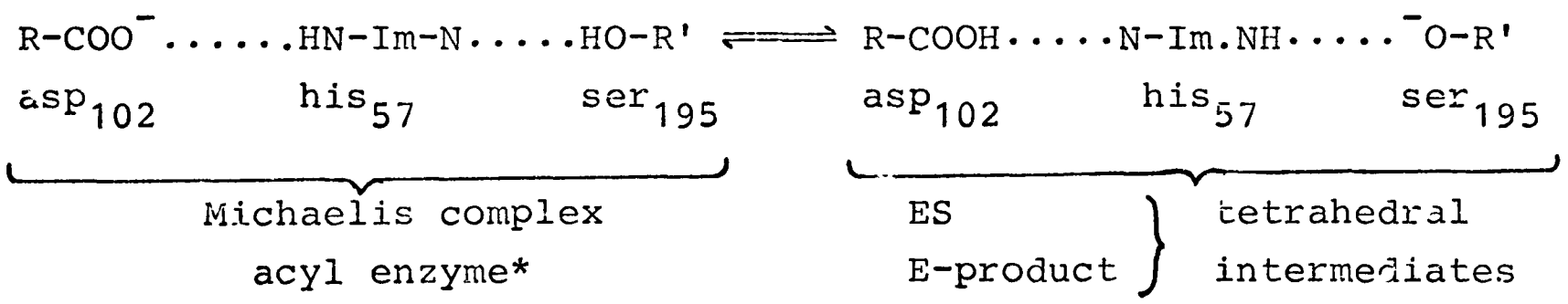
The pH profile of serine protease catalysed reactions is bell-shaped, showing dependency on two groups with pk_{app} of approx. 6.5 - 7.0 and approx. 8.5 (Hammond and Gutfreund, 1955). The α-amino group of ile_{16} has been identified as the group controlling the upper pH limit although in elastase, the substitution of valine for isoleucine gives a much higher pH limit for enzyme activity (pk_{app} = 9.7). In this enzyme, although val_{16} must be buried, its protonation does not appear to be obligatory (Hartley and Shotton, 1971) as enzyme activity is maintained to pH 10.5, and no change in optical rotation occurs until pH 11.0.

(a) Reaction sequence (from Kraut, 1977)



The mechanism is explained in the text. (S), (1), (2) and (3) are the substrate, the Michaelis complex, the tetrahedral intermediate and the acyl enzyme respectively. HX is the leaving group, and HY is the acyl-accepting nucleophile (usually a water molecule). K_S, K_P, k_1-k_4 and k_{-1}-k_{-4} are rate constants.

(b) Proton transfer in the charge relay system



*in the acyl enzyme, H in ser_{195} is replaced by the substrate carbonyl carbon.

Fig. 1.7 Serine protease catalysis

The group with pk_{app} of 6.5 - 7.0 is in the asp_{102} - his_{57} couple of the charge relay system. A unified concept of the responsible group has not been established (Kraut, 1977; Lawson, 1980). Although, it is generally regarded to be the imidazole of his_{57}, an alternative interpretation is that because of the relatively hydrophobic environment of asp_{102} and his_{57}, it is the protonation of asp_{102} which occurs below pH 6.5 - 7.0 (Hunkapiller et al, 1973). The contention concerning the involvement of ser_{195} in the charge relay (section 1.3.1.2) is discussed fully by Kraut (1977) and Lawson (1980).

1.3.2.2 Catalytic mechanism

The catalytic sequence (Fig. 1.7a) involves the interaction of enzyme (E) and substrate (S) to form the Michaelis complex which is converted to a tetrahedral intermediate by covalent bond formation between the substrate carbonyl carbon and the reactive serine O_γ atom. A proton is transferred effectively from the serine hydroxyl group to the leaving group of the substrate. Release of the leaving group results in formation of the acyl enzyme. Deacylation follows by the reaction of the acyl enzyme with a nucleophilic group such as water, and is a reversal of the acylation reaction. Two plausible theories exist to explain the mechanism of proton transfer from and to the reactive serine:-

(1) The charge relay system is the driving force for nucleophilic attack on the substrate carbonyl carbon atom. Proton withdrawal occurs from ser_{195} to his_{57}

(Fig. 1.7b) allowing the formation of the tetrahedral intermediate. On formation of the acyl enzyme, the proton is transferred from his_{57} to the leaving group (X in Fig. 1.7a). A similar proton transfer occurs on deacylation by recruiting a proton from the solvent (HY) (reviewed by Stroud et al, 1975).

(2) Catalysis is possible because the active site is a template for binding the tetrahedral transition state complex (section 1.3.2.2). In the transition state, the carbonyl oxyanion is tetrahedrally skewed and stabilised by hydrogen-bonding. Because the substrate is bound in a strained configuration, nucleophilic attack is possible. The role of the charge relay in this model is binding the proton which in the transition state is being transferred from ser_{195} to the leaving group (shown as being partially attached to both O_γ of ser_{195} and to X, in Fig. 1.7a) (reviewed by Kraut, 1977).

1.3.3 Substrates and inhibitors of serine proteases

1.3.3.1 Substrates

Serine proteases will cleave both polypeptide substrates and small molecules such as synthetic ester and amide derivatives of amino acids. The hydrolysis of such substrates has been used to assay serine protease activity and give information both on the specificity of the reaction catalysed and on the role of residues on either side of P_1 in controlling enzymic activity.

1.3.3.2 Nonprotein inhibitors

Inhibitors which react with the active site serine are used as models of the acyl enzyme intermediate. Such inhibitors include iPr_2P-F, PMSF, and pNPGB. The adducts formed assume tetrahedral geometry, but cannot be cleaved, possibly because of the lack of stabilisation of the oxyanion through hydrogen bonding (Kraut, 1977). However, boronic acid adducts do form hydrogen bonds in a similar manner to normal substrates and are therefore models of the tetrahedral intermediate. Inhibitors such as TPCK and TLCK form a covalent bond with his_{57}.

1.3.3.3 Protein inhibitors

Protein inhibitors of serine proteases have been reviewed recently by Laskowski and Kato (1980). Although they are ubiquitous in nature, their immediate physiological role is not always apparent. Several distinct families occur (Table 1.2) and are thought to have arisen by convergent evolution. The inhibitors range in size from about 5000-8000 daltons for the smallest inhibitors (e.g. Bowman-Birk soybean proteinase inhibitor and potato I inhibitor families) to about 160,000 (inter-α-trypsin inhibitor) Some inhibitors have two or more regions of homology and may have more than one reactive site, often with different specificities.

The mechanism of inhibitory action depends on the existence of a peptide bond within the inhibitor (the reactive site) which combines with the enzyme in a substrate-like manner. Although complex formation occurs readily, bond cleavage at the reactive site is slow and the reaction

Table 1.2 Families of protein inhibitors of serine proteases (from Laskowski and Kato, 1980)

I Bovine pancreatic trypsin inhibitor (Kunitz) family.

II Pancreatic secretory trypsin inhibitor (Kazal) family.

III *Streptomyces* subtilisin inhibitor family.

IV Soybean trypsin inhibitor (Kunitz) family.

V Soybean proteinase inhibitor (Bowman-Birk) family.

VI Potato I inhibitor family.

VII Potato II inhibitor family.

VIII *Ascaris* trypsin inhibitor family.

IX Other families.

does not proceed to completion.

A protease inhibitor of a different type is the plasma protein, α_2-macroglobulin (α_2-M), a molecule of mol. wt. 720,000. The protease is thought to hydrolyse one or more susceptible bonds in α_2-M resulting in a conformational change in α_2-M which traps the enzyme (Barrett and Starkey, 1973). Although covalent bond formation between substrate and inhibitor sometimes occurs (Salveson and Barrett 1980) the active site of the enzyme is not blocked, and hydrolysis of small substrates is still possible.

1.3.4 Concluding remarks

1.3.4.1 Evolution of the serine proteases

The vertebrate serine proteases are thought to have arisen by divergent evolution from an ancestral gene (de Haën et al, 1975) and can be broadly divided into pancreatic (mol. wt. 20,000-30,000) and plasma (mol. wt. generally >60,000) (Fig. 1.8). It is thought that the ancestral gene coded for a protease of about 60,000 daltons, possibly synthesized as an active enzyme. The pancreatic enzymes would have evolved later than the plasma enzymes and the higher mol. wt. of the latter group was possibly a result of selective pressures causing the retention by the renal glomeruli of proteins larger than 55,000 daltons (Windhager, 1969). Conversely, lack of selective pressures in the pancreas resulted in shrinkage of the ancestral gene. However, the existence in plasma of factor D, a protease of

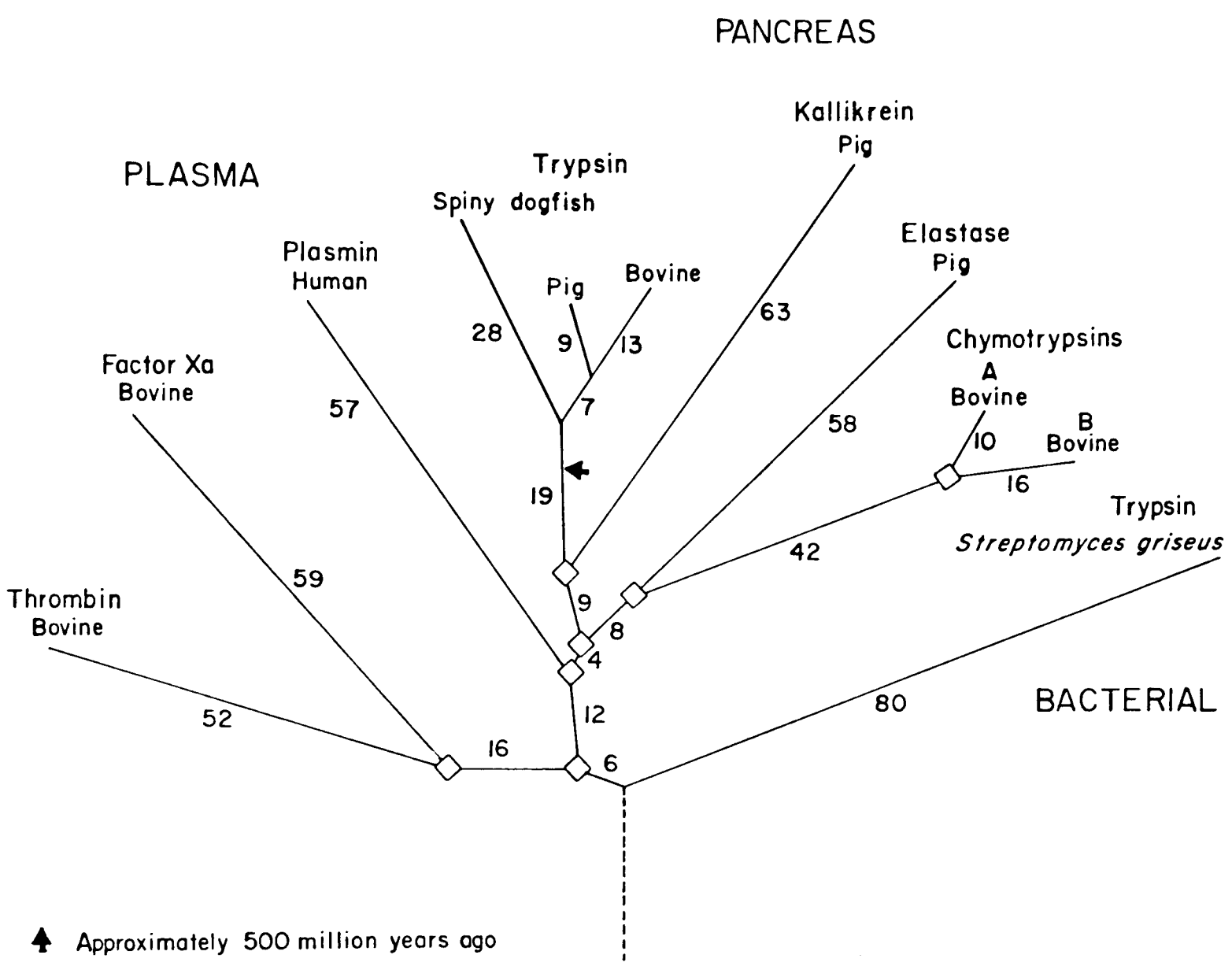


Fig. 1.8 Evolutionary tree of some eukaryote serine proteases and bacterial trypsin (from Dayhoff, 1979)

Diamonds represent gene duplications, and branch lengths are in PAMs (accepted point mutations/100 residues/100 million years). It is estimated that the earliest divergence shown would be approximately 1.3 billion years ago, and the proliferation of the major enzyme types would probably have occurred early in eukaryote evolution.

24,000 daltons, suggests that other selective pressures may also have existed. The higher mol. wt. of blood clotting and fibrinolytic enzymes is accounted for by their "kringle regions". Although their function is unknown, the conservation of amino acid sequence in these regions (Dayhoff, 1979) suggests an important role besides resistance to glomerular filtration. It will be of interest to see if such regions are also present in complement proteases.

Zymogens probably arose very early in the evolution of the serine proteases. Activation of zymogens tends to be by a tryptic-like cleavage (i.e. after an arg or lys residue), and serine proteases with trypsin-like specificity are more widespread than other specificities. In keeping with this, the evolution of proteases with trypsin-like specificity was probably earlier than that of other specificities (Dayhoff, 1979).

1.3.4.2 The serine proteases of the complement system

The serine proteases of the complement system are $C\bar{1}r$, $C\bar{1}s$, factor B, C2, factor $\bar{D}$ and C3b INA. The interesting features of these proteases are

(i) the specificity of $C\bar{1}s$*and factor $\bar{D}$ for cleaving their respective substrates (C2 and factor B) only when these are bound in a Mg^{2+}-dependent complex with another protein (C4b and C3b respectively),

*Although $C\bar{1}s$ will cleave free C2, only C2 cleaved in the C42 complex will form the active C3 convertase (Kerr, 1980).

(ii) the activation mechanism of C1r and the extreme specificity of $\overline{C1r}$ for its substrate C1s, which is only cleaved in the C1 complex,

(iii) the apparent lack of a zymogen for factor $\bar{D}$, and possibly C3b INA,

(iv) the unusual N-terminal structure of the catalytic chains of C2 and factor B which have about 300 residues preceding the conserved isoleucine which is N-terminal in other serine proteases. These two proteases have an extremely short halflife for cleaving their substrate, C3 (at $37^{o}C$, 7 min for $\overline{C42}$ and 15-20 min for $\overline{C3bB}$). However it has been reported that following decay of C2 and factor B from the convertases, although haemolytically inactive, they are still able to cleave synthetic substrates (Cooper, 1971, 1975; Vogt et al, 1977).

It appears then, that the complement serine proteases each contain novel features which will not be fully elucidated until complete structural data are available. Whereas the pancreatic enzymes show broad substrate specificity, the plasma enzymes often cleave only one protein substrate and may have unusually low reactivity towards synthetic substrates and, in the case of C2, factor B and C3b INA, towards well-known inhibitors such as iPr_2P-F. The stabilisation of the active sites of the complement proteases will be particularly interesting, and it is possible that the substrate itself may be important in inducing or stabilising an active conformation.

1.4 FACTOR $\bar{D}$

1.4.1 Introduction

As evidence accumulated that activated factor B was part of a C3-converting enzyme, the search began for a potential activator of factor B. This seemed plausible both by analogy with the classical pathway (Götze and Müller-Eberhard, 1971) and because of the rapid rate at which factor B was converted when it was added to the serum of patients with congenital hypercatabolism of C3. In this disease, both factor B and C3 levels are markedly reduced (Alper and Rosen, 1971). Factor $\bar{D}$ was first demonstrated in human serum by both Rosen and Alper (1972) and Müller-Eberhard and Götze (1972). Its serine protease nature was discovered by Fearon *et al* (1974). Although data has accumulated confirming its role as the physiological activator of factor B, the nature of the physiological control of factor $\bar{D}$ has remained obscure. There has been one report that a precursor form of factor $\bar{D}$ exists (Fearon *et al*, 1974) but this has not been confirmed by others (Lesavre and Müller-Eberhard, 1978).

The study of factor $\bar{D}$ has been hampered by its low plasma concentration (1-2 µg/ml) which has made purification difficult. However, in the last two years, procedures have been described which claim to give homogeneous preparations of human factor $\bar{D}$ (Lesavre *et al*, 1979; Davis *et al*, 1979b). In both these methods, conventional chromatographic procedures were used and the yield of factor $\bar{D}$ based on recovery of haemolytic activity was about

20%. Methods for isolating guinea-pig factor $\bar{D}$ have been described by Dierich et al (1974) and Brade et al (1974a).

Factor $\bar{D}$ haemolytic activity can be measured in several ways. The diffusion method of Martin et al (1976) is based on the reactive lysis of guinea-pig erythrocytes in agarose containing Mg^{2+}, EGTA and factor $\bar{D}$-deficient serum, R(D). The more rapid method of Lesavre et al (1979) uses similar reagents except that the target cells are rabbit erythrocytes and the reaction is carried out in solution. A third method uses EAC43 cells which can be lysed by interaction with B,$\bar{D}$ and a source of C3-9 (Fearon and Austen, 1976).

1.4.2 Physicochemical properties and structural data

Some of the physicochemical properties of human and guinea-pig factor $\bar{D}$ are summarised in Table 1.3. Factor $\bar{D}$ isolated from human plasma is a 3S protein (Müller-Eberhard and Götze, 1972) with a molecular weight of about 24,000 (Volanakis et al, 1977; Lesavre and Müller-Eberhard, 1978; Davis et al, 1979b). Guinea-pig (gp) and human (hu) factor $\bar{D}$ are immunochemically related and in a zymosan-activated system using human complement components, $\bar{D}_{gp}$ could substitute for $\bar{D}_{hu}$ in activating factor B (Brade et al, 1976).

In whole serum, factor $\bar{D}$ has an α-mobility on gel electrophoresis (Martin et al, 1976; Götze and Müller-Eberhard, 1976; Konno et al, 1978). The isoelectric point as determined by isoelectric focussing has been variously reported as 7.4 (Fearon and Austen, 1975a; Götze and

Table 1.3 Physicochemical properties of factor $\bar{D}$[1]

Property	Human factor $\bar{D}$	Guinea-pig factor $\bar{D}$
Concentration in serum	1-2 µg/ml	-
Mol. wt.	24,000	22,000
S-rate	3S	2.6S
Frictional coefficient (f/fo)	-	1.1
Isoelectric point	approx. 7.4	approx. 9.4
Electrophoretic mobility		
in EDTA-serum	α	-
in Ca^{2+}-serum	γ	-
Sensitivity (% of activity remaining)		
Hydrazine	100%[2]	100%
Heat (30 min, 56°C)	75% in serum[3]	30% in serum 5-10% in purified preparations
Trypsin (0.15 mg/ml; 30 min. 37°C)	-	100%(no change in mol. wt.)
Reduction/alkylation	0%[4]	0%
pH	-	pH 4.5, 5% pH 5.5 - 8.0, 100%

[1] Data on guinea-pig factor $\bar{D}$ is from Dierich et al (1974) and Brade et al (1974a). Other references not cited here are given in the text. -, no published data available.

[2] Rosen and Alper (1972).

[3] Martin et al (1976). This observation was corroborated by Volanakis et al (1977).

[4] Volanakis et al (1977).

Müller-Eberhard, 1976) and 8.5 (Martin et al, 1976) while Davis et al (1979b) found that purified factor $\bar{D}$ gave two bands with isoelectric points of 6.6 and 7.0. The isoelectric point of guinea-pig factor $\bar{D}$ is 9.3 - 9.5 (Brade et al, 1974a; Dierich et al, 1974). Although this variation in reported values for the isoelectric point is wide and may reflect differences in the methodology used, they all suggest that the pI is high compared with its electrophoretic mobility. Recent evidence has suggested that the electrophoretic mobility may be modulated by Ca^{2+} or Mg^{2+} ions (Davis et al, 1979b). In the presence of Ca^{2+} ions, both purified factor $\bar{D}$ and factor $\bar{D}$ in whole serum had a γ mobility in agarose gels whereas in the presence of EDTA the mobility was more anodal. This suggested that Ca^{2+} ions may be bound by factor $\bar{D}$. In the presence of EDTA, factor $\bar{D}$ in whole serum had an α mobility while purified factor $\bar{D}$ had a β-γ_1 mobility (Lesavre and Müller-Eberhard, 1978; Davis et al, 1979b). It has been suggested that in the presence of EDTA, factor $\bar{D}$ in whole serum may associate with an as yet unidentified protein (Konno et al, 1978).

The amino acid composition of factor $\bar{D}$ has been reported by two different laboratories (Volanakis et al 1977; Davis et al, 1979b) but the two reports differ significantly in several respects (Table 1.4). Both Lesavre et al (1979) and Davis et al (1979b) have reported that factor $\bar{D}$ stains with periodic acid - Schiff reagent, but extensive carbohydrate studies have not yet been carried out.

Table 1.4 Amino acid composition of human factor D

Amino acid	Residues/100 residues	
	Volanakis *et al* (1977)	Davis *et al* (1979b)
Asp	7.4	9.1
Thr	4.3	4.1
Ser	7.6	7.1
Glu	9.8	9.1
Pro	8.7	7.1
Gly	9.8	10.7
Ala	5.8	10.7
Cys	2.9	-
Val	6.8	8.1
Met	1.4	1.5
Ile	6.0	2.5
Leu	6.9	11.7
Tyr	1.1	2.0
Phe	2.0	0.5
His	7.0	4.1
Lys	7.1	3.6
Arg	5.5	8.1

-, not estimated.

The N-terminal amino acid of factor $\bar{D}$ is isoleucine (Volanakis et al, 1977). Davis et al (1979a, 1979b) reported a tentative sequence of the first nine residues of factor $\bar{D}$ and Volanakis et al (1980) have reported a more extensive but incomplete sequence of the first 57 amino acid residues. The sequence data in both cases show that homology exists between factor $\bar{D}$ and other typical serine proteases.

1.4.3 Enzymic properties

(1) Inhibitors

The serine protease nature of factor $\bar{D}$ is indicated by its inhibition by 5 mM iPr_2P-F (Fearon et al, 1974). The haemolytic activity is also inhibited competitively by 10 mM TAME (Fearon and Austen, 1975b) and noncompetitively by 0.05 mM cyclohexylbutylphosphonofluoridate (Fearon and Austen, 1975b; Götze and Müller-Eberhard, 1976). Dieminger et al (1976) have reported partial inhibition by 0.2 mM PMSF or 20-25 mM ALME or AGLME. However, neither 10 mM TLCK (Fearon et al, 1974) nor trypsin inhibitors from bovine lung, soybean, limabean or ovomucoid (Götze and Müller-Eberhard, 1976) have any effect on haemolytic activity. Factor $\bar{D}$ is also resistant to the natural protease inhibitors in plasma (Lesavre and Müller-Eberhard, 1978; Lachmann, 1979).

(2) Polypeptide substrates

Factor $\bar{D}$ is extremely specific for its physiological substrate, factor B. It splits a single arg-lys bond in factor B, in the sequence

-gln-lys-arg-lys-ile-val-

(Lesavre et al, 1979). In order for cleavage to occur factor B must be in a Mg^{2+}-dependent complex with C3b (Götze and Müller-Eberhard, 1971; Lesavre et al, 1979; Kerr, 1979). This is in contrast to the cleavage of B by trypsin (Vogt et al, 1974; Lesavre and Müller-Eberhard, 1978) or of C2 by $C\bar{1}s$ (Gigli and Austen, 196?a, 1969b) which can both occur without the substrate being complexed. Factor $\bar{D}$ is unable to cleave the proteins casein, haemoglobin, fibrin or elastin (Götze, 1976).

(3) Synthetic substrates

Early studies on the ability of factor $\bar{D}$ to cleave synthetic esters suggested that it had limited esterolytic activity towards a variety of synthetic lysyl and arginyl esters (Volanakis et al, 1977). Dieminger et al (1976) carried out thin layer chromatography on the cleavage products of the synthetic esters, ALME and AGLME, and found that the products formed following cleavage by factor $\bar{D}$ differed in mobility from the products formed following cleavage by trypsin. In later studies, neither Lesavre and Müller-Eberhard (1978) nor Davis et al(1979b) could demonstrate any esterolytic activity toward these two substrates or a variety of others, with the exception of a p-nitroanilide substrate (N-Bz-ile-glu-gly-arg-pNA.HCl [S2222]) which was weakly hydrolysed. It is probable, therefore, that the cleavage of synthetic substrates observed in earlier studies was caused by a contaminating protease and not by factor $\bar{D}$. However, ALME and AGLME

(Dieminger et al, 1976) and certain synthetic peptides which resemble the primary structure of the cleavage site in factor B (Lesavre et al, 1980) do inhibit B cleavage by factor $\bar{D}$, suggesting that binding of these substrates in the active site is occuring.

1.4.4 Physiological role of factor $\bar{D}$

Factor $\bar{D}$ functions in vivo as the activator of the alternative pathway C3 convertase, $\overline{C3bB}$, by cleaving factor B at an arg-lys bond (Lesavre et al, 1979). Although the formation of the C3bB complex requires Mg^{2+} ions, activation by factor $\bar{D}$ is independent of cofactors (Vogt et al, 1975; Lesavre and Müller-Eberhard, 1978) and can proceed in the presence of EDTA. The function of factor $\bar{D}$ appears to be solely enzymic, although some conflicting observations still require clarification.

(1) Binding of factor $\bar{D}$ to the convertase or to the activating surface

Lesavre and Müller-Eberhard (1978) observed no consumption of factor $\bar{D}$ as a result of alternative pathway activation either in serum or on interaction of isolated factor $\bar{D}$ with the cellbound C3 convertase. No depletion of factor $\bar{D}$ from the fluid phase was observed, nor could ^{125}I-labelled factor $\bar{D}$ be recovered from red cell pellets or Sepharose columns containing surface-bound C3 convertase. Zymosan-depletion of alternative pathway components did not affect factor $\bar{D}$ haemolytic activity or antigen concentration, and sucrose-density gradient studies on the fluid-phase convertase using ^{125}I-labelled factor $\bar{D}$

did not reveal any association with $\overline{C3bB}$. Brade et al (1974b) using a zymosan-bound convertase, and Vogt et al (1974) using $\overline{CVF,B}$, have shown that other proteolytic enzymes can replace factor $\bar{D}$ in the activation of factor B. Both Brade et al (1974a) using zymosan, and Fujita et al (1977) using immune aggregates have noted the requirement for both factors B and $\bar{D}$ for the regeneration of decayed convertase activity. However, Brade et al (1974a) found that although factor $\bar{D}$ appeared to bind to zymosan, this effect was favoured by low ionic strength (0.077) and unphysiological pH (6.0). The ability of factor $\bar{D}$ to bind to a variety of macromolecules at low ionic strength has also been observed by Dierich et al (1974).

However, Gadd and Reid (1981a) incubated immune aggregates with serum at physiological pH and ionic strength, and found consumption of up to 50% of factor $\bar{D}$ activity. Furthermore, Dierich et al (1974) noted a complex of CVF,B and $\bar{D}$ by gel filtration on Sephadex G200. However, although they used physiological pH (7.5), the ionic strength was not clearly indicated.

A possible interpretation of the conflicting data is that under physiological conditions, factor $\bar{D}$ may bind transiently to the activating surface, but its primary role is enzymatic and it is not an essential subunit of the convertase.

(2) C3 cleavage by the unactivated convertase

Another confusing issue has been the observation that factor B appears to have limited C3-cleaving ability without prior activation by factor $\bar{D}$, even if both factor B and C3 are pretreated with iPr_2P-F (Vogt et al, 1977). Daha et al (1976) using a C3NeF-stabilised convertase, and Fearon and Austen (1975a) using a properdin-stabilised convertase, made similar observations, as have Day et al (1976) and Brade et al (1977a). However, other studies have shown an absolute requirement for factor $\bar{D}$ (Schreiber et al, 1976; Brade et al, 1977a; Lesavre and Müller-Eberhard, 1978). It is possible that on binding to C3b, a conformational change in factor B exposes the active site of the enzyme so that even although it may be structurally distorted, the site is able in a limited way to interact with its substrate, C3.

(3) Regulation of factor $\bar{D}$ activity

Because factor $\bar{D}$ will not cleave factor B unless it is bound to C3b, the activity of factor $\bar{D}$ will effectively be controlled by circulating C3b levels. Regulation will be discussed further in Chapter 6.

1.4.5 Miscellaneous observations

It has been observed that factor $\bar{D}$ will competitively inhibit thrombin-induced platelet aggregation in a nonenzymatic manner (Davis and Kenney, 1979). While the affinity for the thrombin receptor on platelets may merely reflect structural similarities between the two proteins, Davis and Kenney have suggested that this may

be of significance in physiological complement - platelet interactions.

Davis et al (1979b) have found that factor $\bar{D}$ binds to heparin - Sepharose. Since certain cell-surfaces contain heparan sulphate which is structurally similar to heparin (Kazatchkine et al, 1979) it would be interesting to discover if factor $\bar{D}$ showed any binding affinity for these surfaces under physiological conditions.

1.4.6 Genetics and biosynthesis

Little data exists on the genetics of factor $\bar{D}$ but it is not thought to be linked to any of the three known loci for human complement components (C4/C2/factor B; C6/C7; and C3) (Lachmann and Hobart, 1979). Isoelectric focussing of purified factor D has shown two protein bands containing factor $\bar{D}$ haemolytic activity, a major band with an isoelectric point at pH 7.0 and a minor one at pH 6.6 (Davis et al, 1979b) suggesting that at least two allotypes may exist. In two earlier population studies, only one major variant was found except for the rare occurrence (0.4%) of a second variant which occurred in individuals of West African parentage (Martin et al, 1976; Hobart and Lachmann, 1976). No deficiencies of factor $\bar{D}$ have been described.

Functionally active factor $\bar{D}$ is synthesized and secreted by human peripheral blood monocytes (Whaley, 1980; de Ceulaer et al, 1980) and by human synovial and guinea-pig peritoneal macrophages (Bentley et al, 1978; de Ceulaer et al,

1980). Bentley et al (1976, 1977) have not found factor $\bar{D}$ activity in culture supernatants from mouse macrophages but it was unclear from their experiments whether this was because the protein was not synthesized by these cells or whether some other factor prevented the detection of factor $\bar{D}$ activity.

1.5 AIMS OF THE PROJECT

1.5.1 Purification and characterisation of factor $\bar{D}$

When the work described here was begun, a purification method which gave a sufficiently pure preparation for both enzymic studies and extensive sequencer studies was not available. The first task, therefore, was to develop such a method and to characterise the purified enzyme.

1.5.2 The primary structure of factor $\bar{D}$

The serine protease nature of factor $\bar{D}$ had been described by Fearon *et al* (1974) and in 1977, Volanakis *et al* identified the N-terminal amino acid (isoleucine). Therefore, the elucidation of the primary structure, which would confirm its role as a serine protease, was undertaken as the major part of this thesis.

1.5.3 Enzymic activity of factor $\bar{D}$

Published data on factor $\bar{D}$ had shown little evidence for a zymogen form of the enzyme, except for one report by Fearon *et al* (1974). Furthermore, data on its nonhaemolytic enzymic activity were conflicting but had suggested that factor $\bar{D}$ was highly specific for its physiological substrate, factor B, and only when this was bound to C3b. The purpose of the work described here was to repeat the experiments of Fearon *et al* (1974) and to further investigate the substrate specificity of factor $\bar{D}$ and its sensitivity to inhibitors.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals

General chemicals and organic solvents, including acrylamide and N,N'-methylene-bis-acrylamide were obtained from Fisons Scientific Apparatus, Loughborough, U.K.

Dithiothreitol, iodoacetic acid, iodoacetamide, guanidine hydrochloride, 1-dimethylnaphthalene-5-sulphonyl chloride (dansyl chloride), standard dansyl amino acids, soluble starch, TEMED (N,N,N',N'-tetramethylethylene diamine), periodic acid and agarose were from BDH Chemicals Ltd., Poole, Dorset, U.K.

Ethylene-glycol-bis-(β-amino-ethyl ether) N,N'-tetra-acetic acid (EGTA), 1-0-methyl-α-D-glucopyranoside, Tris (hydroxymethyl) aminomethane, Hepes (N-2-hydroxymethyl-piperazine-N'-2-ethane sulphonic acid) were from Sigma Chemical Co., London, U.K.

Mercaptoethane-sulphonic acid, cyanogen bromide, ninhydrin, o-iodosobenzoic acid were from Pierce Chemical Co., Rockford, Illinois, U.S.A.

PPO (2,5-diphenyloxazole), POPOP (1,4-di-[2-(4-methyl-5-phenyloxazolyl)]-benzene), were from Koch-Light Laboratories Ltd., Colnbrook, Bucks, U.K.

Iodo-[2-^{14}C]-acetic acid (54 mCi/mmole) and NCS tissue solubiliser were supplied by Radiochemical Centre,

Amersham, Bucks., U.K.

Standard amino acids were from Roche, Welwyn Garden City, U.K. or Sigma or Fisons.

Di-[1,3,-^{14}C]-isopropyl phosphofluoridate (100 μCi/μmole) was purchased from New England Nuclear.

Schiff's (Feulgen) stain and ion agar were from Difco Labs., Surrey, U.K.

Quadrol (N,N,N',N'-tetrakis-(2-hydroxypropyl)-ethylene-diamine-trifluoroacetate) was from Beckman Instruments, Palo Alto, California, U.S.A. Polybrene (1,5-dimethyl-1, 5-diazo-undecamethylene polymethobromide) was from Aldrich Chemical Co., R.N. Emanuel, Wembley, Middlesex, U.K.

2.1.2 Enzymes and Marker Proteins

Collagenase (Type vi, from C. histolytica), ovalbumin, bovine α-chymotrypsinogen A, whale skeletal muscle myoglobin, eggwhite lysozyme, horse heart cytochrome C, and bovine insulin were all obtained from Sigma Chemical Co., London, U.K. Bovine serum albumin was obtained from Armour Pharmaceutical Co., Eastbourne, Sussex, U.K.

Trypsin treated with TPCK to inactivate any chymotrypsin contaminant was obtained from Worthington Biochemical Corporation, Freehold, New Jersey, U.S.A.

Protease from Staphylococcus aureus V8, was supplied by Miles Laboratories, Stoke Poges, U.K.

Carboxypeptidase-Y was obtained from United Breweries Ltd., Copenhagen, Denmark.

2.1.3 Protease inhibitors, synthetic substrates and antisera

The following were obtained from Sigma Chemical Co., London, U.K.: soyabean trypsin inhibitor, type 1-S: lima bean trypsin inhibitor, type 11-L; aprotinin from bovine lung; pepstatin; leupeptin hemisulphate; ovomucoid, type 11-0 (partially purified and containing ovoinhibitor) from chicken eggwhite; ovomucoid, type 111-0 (free of ovo-inhibitor) from chicken eggwhite; N-tosyl-L-arginine methylester (TAME); N-benzoyl-L-arginine ethyl ester (BAEE); N-acetyl-1-lysine methyl ester (ALME); N-acetyl-glycyl-L-lysine methyl ester (AGLME).

Diisopropyl phosphofluoridate (iPr_2P-F) was from Aldrich Chemical Co., Milwaukee, Wisconsin, U.S.A.

p-Nitrophenyl-p'-guanidinobenzoate was from Vega Fox Biochemicals, Tucson, Arizona, U.S.A.

p-Nitranilide substrates were obtained from KABI, Mölndal, Sweden.

The following were kindly provided by members of the MRC Immunochemistry Unit: $C\bar{1}$-inhibitor and $\alpha_2$macroglobulin (Dr. R.B. Sim); inter-α- trypsin inhibitor (Miss E.M. Press); BDC-OH (4,4'bis-dimethyldiphenylcarbinol) (Dr. J.H. Harrison); anti-$C\bar{1}$s $F(ab')_2$ and anti C1q antisera (Dr. K.B.M. Reid); thiobenzyl-benzoxycarbonyl-lysinate (ZLSB) (Prof. R.R. Porter).

2.1.4 Media for column chromatography

QAE-Sephadex A-50; CM-Sephadex C-50; Sephadex G-75; Sephadex G-75 (superfine); Sephadex G-50 (superfine); Concanavalin A-Sepharose; and Agarose-lysine were all obtained

from Pharmacia Fine Chemicals, Uppsala, Sweden. Ion-exchange celluloses DE-32 and CM-32 were from Whatman Ltd., Maidstone, Kent, U.K.

2.1.5 Plasma and erythrocytes

2.1.5.1 Plasma and serum for the isolation of factor $\bar{D}$

Fresh or outdated plasma was used as a source of factor $\bar{D}$. Plasma, collected in sodium citrate anticoagulant, was obtained from the Churchill Hospital, Oxford. When serum was required the plasma was clotted overnight at 4°C by the addition of 1.0 M $CaCl_2$ to give a final concentration of 20 mM. The clot was removed by centrifugation and filtration through muslin, and the serum was stored at -20°C.

2.1.5.2 Factor $\bar{D}$-deficient serum

Fresh frozen plasma was clotted as described above, and stored at -70°C. Serum was depleted of factor $\bar{D}$ by fractionation on a column of Sephadex G-75 equilibrated in 0.1 M Tris/HCl/0.2 M NaCl/2 mM EDTA, pH 8.0 (at 4°C), as described by Martin et al (1976).

2.1.5.3 Guinea-pig erythrocytes

Fresh guinea-pig blood obtained by cardiac puncture and collected into sodium citrate anticoagulant was supplied by University Park Farm, Eynsham. Erythrocytes were prepared as described by Martin et al (1976).

2.2 METHODS

2.2.1 General Methods

2.2.1.1 Chromatographic procedures

Gel filtration media was prepared and used as advised by the manufacturers. Apparent molecular weights of peptides purified by gel filtration were estimated by comparison of the partition coefficient, K_{av}, of the peptides with those of marker proteins, on a plot of K_{av} versus log. molecular weight. K_{av} is calculated using the relationship

$$K_{av} = \frac{Ve - Vo}{Vt - Vo}$$

where Ve, Vt and Vo are respectively, the elution volume of the sample, the total column volume and the void column volume.

Ion exchange media were prepared, run and equilibrated according to the recommendations of the manufacturer, using stepwise or linear gradients. The columns were pumped at a constant pressure using a peristaltic pump (Gilson Minipuls 2 or Pharmacia P--3).

Affinity columns and most gel filtration columns were run under gravity using downward flow. For large gel filtration columns, upward flow was used.

2.2.1.2 Estimation of protein concentration

The concentration of protein in column eluates was monitored by absorbance at 280, 230 or 215 nm. Estimates of protein yields during purification were obtained from A_{280}. The final yields of factor $\bar{D}$ following purification were

determined from the leucine content obtained by amino-acid analysis after 48 h acid hydrolysis at 110°C.

2.2.1.3 Buffer solutions and conductivity measurements

The compositions of buffer solutions were calculated according to tables published by Dawson et al (1978). Except where specified, pH was measured at 25°C. The conductivity of buffer solutions and column fractions was measured at 4°C using a Radiometer CDM3 conductivity meter fitted with a CDC 114 electrode. Conductivity was expressed in mS.

2.2.1.4 Ultrafiltration

Protein solutions were concentrated at 4°C by ultrafiltration using Amicon Diaflo ultrafiltration units (Amicon BV, Oosterhout, Holland). For the concentration of crude protein solutions, Amicon PM10 membranes were used. Amicon UM2 membranes were used for solutions containing molecular species of mainly 25,000 daltons or less.

2.2.1.5 Dialysis and lyophilisation

Dialysis was carried out at 4°C. Freeze-drying was carried out under reduced pressure (≤ 150 millitorr) using an FTS Systems freeze-drier.

2.2.1.6 Detection of radioactivity

β-emission from ^{14}C-labelled samples was counted in an LKB-Wallac 1210 Ultrobeta counter.

(i) Polyacrylamide gel slices Gel slices (1.2-1.5 mm) were placed in glass scintillation vials and incubated for 2 h at 50°C in 1.0 ml distilled water: NCS tissue solubiliser (1:9 by volume). After cooling, 10 ml of toluene containing 0.5% (w/v) PPO and 0.03% (w/v) POPOP was added to each vial, and the radioactivity counted for 5 min.

(ii) Aqueous solutions Samples (10-30 µl) were counted for 5 min in 10 ml 1,4-dioxan containing 2% (w/v) naphthalene and 0.5% (w/v) PPO.

The efficiency of counting of iodo-[2-^{14}C]-acetic acid was 84 ± 2% in the dioxan-based scintillant and 66 ± 5% in the toluene-based scintillant.

2.2.2 Electrophoresis

2.2.2.1 Polyacrylamide gel electrophoresis

Electrophoresis was carried out either as described by Laemmli (1970) using 15% or 20% (w/v) polyacrylamide slab gels overlaid with 3% (w/v) polyacrylamide stacking gel, or as described by Fairbanks et al (1971) using 7% or 12% (w/v) polyacrylamide cylindrical gels. All gels contained 1% (w/v) sodium dodecyl sulphate (SDS). Running buffer for slab gels contained 0.1% (w/v) sodium dodecyl sulphate (SDS) and for cylindrical gels, 0.2% (w/v) SDS.

Non-reduced protein samples (5-10 µg for slab gels, 20-40 µg for cylindrical gels) in 0.1 M Tris/HCl/4 M Urea/ 1% (w/v) SDS, pH 8.0, were prepared by incubation in 20 mM iodoacetamide for 2 min at 100°C. Reduced samples in the same buffer were prepared by incubation in 20 mM dithiothreitol for 2 min at 100°C, followed by alkylation in 40 mM iodoacetamide for 2 min at 100°C.

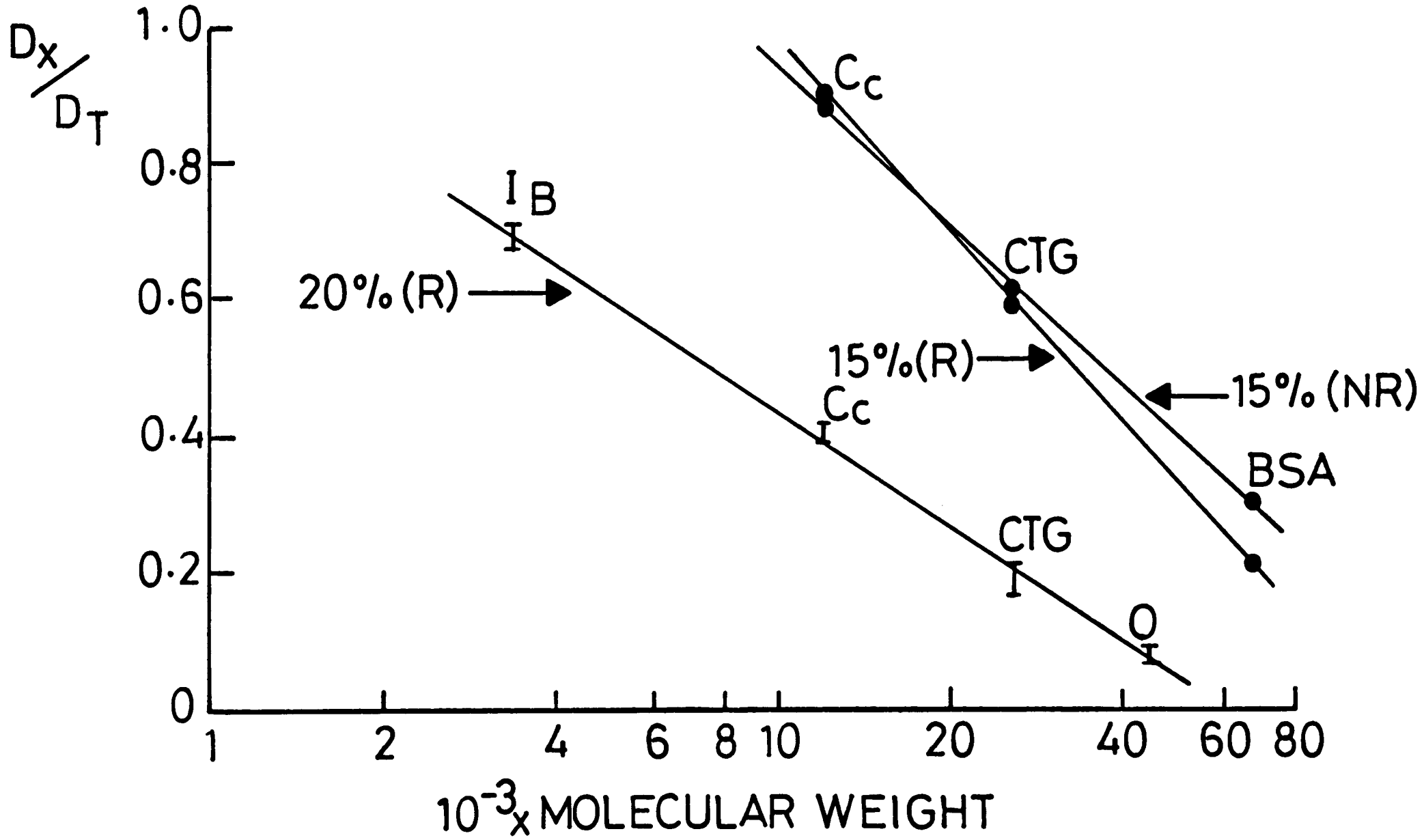
Gels were stained for protein using Coomassie Brilliant Blue (Weber and Osborn, 1969) or for carbohydrate using periodic acid-Schiff stain (Kapitany and Zebrowsky, 1973). Mobility was determined relative to bromophenol blue. Apparent molecular weights of proteins and peptides were determined by comparison with reduced and non-reduced marker proteins as described by Weber and Osborn, 1969 (molecular weights in brackets): bovine serum albumin (67,000), ovalbumin (45,000), bovine pancreas α-chymotrypsinogen A (25,700), horse heart cytochrome C (12,400), bovine insulin B chain (3,400). The line of "best-fit" was computed by least squares regression and the molecular weights of unknowns were calculated accordingly. Typical plots are shown in Fig. 2.1.

Protein was eluted from gels using a modification of the method of Volanakis et al (1980). The protein sample (10 µg) was loaded in 25% (v/v) glycerol, without prior reduction or alkylation. Following electrophoresis, the gel was sliced, the slices pulverised with a spatula and the protein eluted into 0.05 M N-ethylmorpholine/acetic acid/0.001% (w/v) SDS, pH 8.5, by gently rocking for 48 h at 4°C. Eluates were dialysed extensively against distilled water and freeze-dried. The eluates were resuspended in 20 µl 0.004 M PBS, pH 7.2.

Gels were scanned in a Gelman DCD-16 digital computing densitometer. Coomassie Blue was detected at 560 nm and PAS-stain at 525 nm.

Fig. 2.1 Standard plots for molecular weight estimation by polyacrylamide gel electrophoresis

D_X/D_T is the mobility of the standard protein (D_X) relative to that of the bromophenol blue marker (D_T). Markers used (with abbreviation and mol. wt. in brackets) were bovine serum albumin (BSA, 67,000), ovalbumin (O, 45,000), chymotrypsinogen (CTG, 25,700), cytochrome c (C_C, 12,400) and insulin B chain (I_B, 3,400). SDS gels containing 15% or 20% polyacrylamide were used. R, marker proteins were reduced and alkylated; NR, marker proteins were alkylated without reduction.


D_x/D_T
1·0
0·8
0·6
0·4
0·2
0
I_B
20% (R)
C_C
CTG
15% (R)
C_C
15% (NR)
BSA
CTG
O
1
2
4
6
8
10
20
40
60
80
10^{-3} x MOLECULAR WEIGHT

2.2.2.2 Analytical high voltage electrophoresis at pH 1.9

To detect low molecular weight peptides and free amino-acids, high voltage electrophoresis (Katz et al, 1959) was used. Samples (10-20 nmoles/cm) were applied in a narrow strip to Whatman No.1 chromatography paper. A marker solution containing ε-DNP lysine, lysine, arginine and methyl green was run in parallel together with other amino-acids relevant to the unknown being investigated. The chromatogram was developed in a pH 1.9 buffer made up of formic acid : acetic acid : water (2:8:90 by volume), using an electrical field of 3.1 kV.

Following electrophoresis, strips from the edge of each sample on the chromatogram were stained to locate the position of peptides and amino acids. The stains used were:-

(i) Ninhydrin/pyridine 0.2% (w/v) ninhydrin in acetone containing one drop of pyridine. Strips were dipped in the stain and allowed to dry for 2 h at 25°C, then developed for a further 15 min at 80°C.

(ii) Starch-iodide

Solution A : 0.24% (v/v) aqueous sodium hypochlorite

Solution B : 95% (v/v) aqueous ethanol

Solution C : 0.5% (w/v) starch/0.5% (w/v) potassium iodide in distilled water

Strips were sprayed successively with each solution, allowing at least 15 min drying period at 25°C between each spray.

Peptides and amino acids detected by staining, were eluted from the chromatogram using 0.1 M ammonia, using the method of Edstrom (1968). Aliquots (usually 50% of the total elution volume) were hydrolysed with 6 M HCl for amino acid analysis.

2.2.3 Haemolytic activity of factor $\bar{D}$ and factor B

The haemolytic activities of factors $\bar{D}$ and B were assayed using the haemolytic diffusion plate assay of Martin et al (1976). Reagents were prepared as described by Martin et al (1976) except for the modification in the preparation of R(D) reagent described in section 2.1.5. The assay for factor $\bar{D}$ was standardised using dilutions of a pooled normal human serum. One unit of factor $\bar{D}$ activity was defined as the haemolytic area obtained using 10 µl of the standard serum. Using this technique, 1 ng of factor $\bar{D}$ could be detected with confidence. This is roughly equivalent to a 1/10 dilution of normal human serum. The standard deviation, σ, for the haemolytic area using 10 µl standard serum was 7.5 mm^2 (sample size = 36) and therefore the 2σ limit was 15 mm^2. This means that for any one sample, a variation in the calculated haemolytic area of $\pm$ 15 mm^2 can be expected. To minimize experimental error, the concentration of samples for assay was kept, where possible, within the range 0.1 - 1.0 units per 10 µl. An indication of the experimental error is given in Fig. 5.1(b).

2.2.4 Isolation of factor $\bar{D}$

2.2.4.1 Isolation of "precursor factor $\bar{D}$"

Plasma was fractionated as described by Fearon *et al* (1974). Plasma (85 ml) was dialysed against two changes of 3 litres 0.0035 M sodium phosphate, pH 8.0 over a period of 48 h. After removal of the euglobulin by centrifugation for 1 h at 2500 g, the pseudoglobulin was loaded onto a column (5 cm x 90 cm) of QAE-Sephadex A-50 equilibrated with 0.0035 M sodium phosphate, pH 8.0. The column was washed with the same buffer at 60 ml/h until the A_{280} returned to zero, and then a linear gradient was developed using 100 ml 0.0035 M sodium phosphate, pH 8.0 and 1000 ml of the same buffer containing 0.3 M sodium chloride.

2.2.4.2 Preliminary investigations for a purification method for factor $\bar{D}$

Euglobulin precipitation at different pH values, ammonium sulphate fractionation of serum, addition of inhibitors and preparations from both plasma and serum were all evaluated in the initial stages of the work described here for the purification of factor $\bar{D}$. In all cases, factor $\bar{D}$ haemolytic activity was assayed by the method of Martin *et al* (1976) (section 2.2.3).

(i) Euglobulin precipitations Samples of serum (5 ml) were dialysed for 20 h against 200 ml 0.02 M sodium acetate/ acetic acid, pH 4.0 or pH 4.5 or against 200 ml 0.02 M sodium phosphate/5 mM EDTA over a pH range from 5.0 to 8.0. The dialysed samples were centrifuged in a Beckman L5 - 65

ultracentrifuge at 70,000 g for 30 min. The euglobulin precipitates were redissolved in 5 ml 0.004 M sodium phosphate/0.15 M NaCl, pH 7.2.

(ii) Ammonium sulphate fractionation Serum (9 ml) was fractionated using saturated ammonium sulphate (Chase and Williams, 1968). Fractions were made at 10, 20, 30, 40 and 50 percent saturation with respect to ammonium sulphate. The precipitates were centrifugated at 3000 g for 60 min and redissolved in 0.004 M sodium phosphate/ 0.15 M NaCl, pH 7.2.

(iii) Inhibition of plasma proteases Before precipitation of euglobulins, serum was treated with 2.5 M iPr_2P-F in propan-2-ol to give a final concentration of 2.5 mM, and with benzamidine hydrochloride to give a final concentration of 0.5% (w/v) (Kerr, 1979). These concentrations were maintained in the initial stages of purification by addition of the inhibitors to dialysis fluid and buffers for CM-Sephadex C-50 fractionation.

(iv) Preparations from plasma Plasma for factor $\bar{D}$ purification was pretreated as follows, to prevent the activation of Hageman factor and plasma proteases:-

Contact with glass was avoided in all manipulations in the initial stages of purification. Erythrocytes were removed by centrifugation at 2000 g for 30 min. To the centrifuged plasma was added ε-aminocaproic acid to give a final concentration of 25 mM and 1 M $BaCl_2$ to give a final concentration of 40 mM. The mixture was stirred for 30 min at 4°C and the bulky precipitate which formed,

was removed by centrifugation at 10,000 g for 20 min. To the supernatant polybrene was added to give a final concentration of 0.1% (w/v). The supernatant was then dialysed for 20 h against 5 mM EDTA, pH 5.4 (10 litres/ 1 litre plasma) containing 25 mM ε-aminocaproic acid. Purification was then carried out, using the routine method described below (section 2.2.4.3) except that 5 mM EDTA was added to all buffers used for the CM-Sephadex C-50 column.

(v) Preparations from serum Serum was purified as described in section 2.2.4.3.

2.2.4.3 Routine purification of factor $\bar{D}$

The method which follows was that which was found to give the purest preparation of factor $\bar{D}$ as assessed by specific activity (haemolytic activity/mg), homogeneity on reduced SDS/polyacrylamide gels, and N-terminal amino acid sequence. Other methods which were initially tried are discussed in section 3.2.

Serum (4000 ml) was dialysed for 20 h at 4°C against 40 litres of 5 mM EDTA, pH 5.4. The euglobulin precipitate was removed by centrifugation at 2000 g for 1 h at 4°C. The conductivity of the pseudoglobulin was adjusted to 12-14 mS using 1 M sodium phosphate buffer, pH 6.0 and loaded onto a column (18 cm x 8 cm) of CM-Sephadex C-50, equilibrated with 0.2 M sodium phosphate, pH 6.0. The column was then washed with 1500 ml of 0.2 M sodium phosphate, pH 6.0 and the factor $\bar{D}$ was eluted using 2000 ml of 0.4 M sodium phosphate buffer pH 6.0. The flow rate was 600 ml/h during

the 0.2 M sodium phosphate wash, and 50 ml/h during the elution of factor $\bar{D}$.

The fractions containing factor $\bar{D}$ haemolytic activity were pooled, and $(NH_4)_2SO_4$ was added to 50% saturation (291 g/litre). The suspension was stirred for 2 h at 4°C, then centrifuged at 10,000 g for 2 h. $(NH_4)_2SO_4$ was added to the supernatant (125 g/litre) to give 70% saturation and the suspension was stirred for 2 h at room temperature. The precipitate was dissolved in 40 ml of 0.1 M Tris/HCl/0.2 M NaCl/2 mM EDTA, pH 8.0 (at 4°C) and, after centrifugation, applied to a column (5 cm x 100 cm) of Sephadex G-75 equilibrated with the same buffer. The column was run at a flow rate of 30 ml/ h. The fractions containing factor $\bar{D}$ haemolytic activity were pooled and concentrated to 10 ml using an Amicon Diaflo Ultrafiltration cell fitted with a UM2 membrane.

The material from the Sephadex G-75 step from two different preparations was pooled and dialysed against two changes of 500 ml of 0.23 M acetic acid/NaOH buffer, pH 5.2 and applied to a column (1.5 cm x 20 cm) of CM-Cellulose 32 equilibrated with the same buffer. The column was washed with the starting buffer at 30 ml/h until the A_{230} of the eluate was zero, and then a linear gradient was developed using 120 ml of 0.23 M acetic acid/ NaOH buffer, pH 5.2 and 120 ml of the same buffer containing 0.3 M NaCl. The fractions containing factor $\bar{D}$ haemolytic activity were pooled and concentrated to 10 ml, then dialysed against 0.01 M Tris/HCl/0.15 M NaCl, pH 8.0, containing 1 mM $MgCl_2$ and 1 mM $CaCl_2$. After dialysis the sample (10 ml) was applied to a column (1 cm x 12 cm) of

Concanavalin A-Sepharose, equilibrated with the same buffer, and eluted at a flow rate of 10 ml/h. The factor $\bar{D}$ activity was not retarded on the column. A contaminant of the same apparent molecular weight as factor $\bar{D}$ in non-reducing conditions on SDS/polyacrylamide gels, was bound to the column and could be eluted with 2.5% (w/v) 1-O-methyl-α-D-glucopyranoside dissolved in the same buffer used to equilibrate the column.

The factor $\bar{D}$ obtained from the Concanavalin A-Sepharose column was dialysed extensively against 0.5 M acetic acid and then freeze-dried prior to use in sequencing studies.

2.2.5 Immune diffusion

Ouchterlony immune diffusion was carried out in petri dishes as described by Mayer (1961). Gels containing 1.5% (w/v) agar in 0.2 M Tris/HCl/0.01% (w/v) thiomersol, pH 8.0, or 1% (w/v) agarose in 0.05 M Tris/HCl/glycine/0.15 M NaCl/ 10 mM EDTA/0.01% (w/v) sodium azide, pH 8.0 were used. Wells (3 mm diameter) were cut, and after addition of samples the plates were incubated for 2-3 days at 37°C. Plates were washed extensively with saline and stained briefly with Coomassie Blue (section 2.2.2.1).

2.2.6 Spectrophotometric assays with synthetic substrates

Hydrolysis of synthetic substrates was followed spectrophotometrically using a Beckman Model 35 Spectrophotometer attached to a Beckman Recorder/Controller. Samples were assayed in cuvettes maintained at 37°C in a temperature controlled compartment. The reference cell contained substrate and buffer alone. For samples showing

insignificant cleavage of a substrate, the absorbance was followed for 20-60 min using the recorder, or was read at defined intervals between 0 h and 5 h on the spectrophotometer.

The general method for the assays was as follows: substrate was added to the assay buffer in both a reference cell and a test cell. In addition, enzyme buffer was added to the reference cell. The solutions were mixed and the cells allowed to equilibrate at 37°C. The hydrolysis of substrate was initiated by addition of enzyme to the test cell, and after rapidly mixing, the change in absorbance was recorded.

The activity with respect to a given chromogenic substrate was calculated as μmole substrate hydrolysed/ min/mg enzyme, using the general formula

$$\frac{\dfrac{\Delta A/\text{min}}{\varepsilon_M \times 10^{-6}} \times \dfrac{\text{reaction volume}}{10^3}}{\text{mg enzyme}}$$

where ΔA is the change in absorbance and ε_M is the molar extinction coefficient ($M^{-1}cm^{-1}$) for the product of substrate hydroysis.

2.2.6.1 Nitroanilide substrates

The substrates used are listed in Table 2.1.

Assay conditions: substrate (1 mM in distilled water) 100 μl
enzyme 10 - 200 μl
assay buffer (0.1 M Hepes/0.05 M NaCl, pH 8.0) to 1.0 ml.

ε_M (405 nm) (p-nitroaniline) = 11300

Table 2.1 p-nitroanilide substrates

Substrate No.	Structure
S-2393	H-D-Ile-Gly-Arg-pNA·2HCl
S-2322	H-D-Val-Gly-Arg-pNA·2HCl
S-2444	pyro-Glu-Gly-Arg-pNA·HCl
S-2222	Bz-Ile-Glu-Gly-Arg-pNA·HCl
S-2224	Bz-Phe-Gly-Arg-pNA·HCl
S-2343	Bz-Leu-Ala-Arg-pNA·HCl
S-2344	H-Leu-Ala-Arg-pNA·2HCl
S-2347	Bz-Val-Ala-Arg-pNA·HCl
S-2407	pyro-Glu-Ala-Arg-pNA·HCl
S-2345	Bz-Ile-Ala-Arg-pNA·HCl
S-2160	Bz-Phe-Val-Arg-pNA·HCl
S-2238	H-D-Phe-Pip-Arg-pNA·2HCl
S-2302	H-D-Pro-Phe-Arg-pNA·2HCl
S-2266	H-D-Val-Leu-Arg-pNA·2HCl

The molar extinction coefficient was determined (i) from the absorption spectrum of a substrate sample completely hydrolysed enzymatically and (ii) from the A_{405} of 0.01 M p-nitroaniline prepared in the assay buffer.

2.2.6.2 Arginyl esters

(i) TAME (Walsh, 1970)

Assay conditions: TAME (0.01 M in distilled water) 150 µl
enzyme 50 - 200 µl
assay buffer (0.046 M Tris/HCl/0.0115 M $CaCl_2$, pH 8.1) to 1.5 ml

ε_M (247 nm) (arginyl toluene sulphonic acid) = 409

(ii) BAEE (Schleuning and Fritz, 1976)

Assay conditions: BAEE (0.005 M in 0.2 M Tris/HCl, pH 8.0) 150 µl
enzyme 50 - 150 µl
assay buffer (0.2 M Tris/HCl, pH 8.0) to 1.5 ml.

ε_M (253 nm) (arginyl benzoic acid) = 1150

2.2.6.3 ZLSB (Green and Shaw, 1979)

Assay conditions: ZLSB (1 mM in distilled water 10 µl
DTNB (0.13 M in dimethyl-formamide) 5 µl
enzyme 10 - 50 µl
assay buffer (0.1 M Tris/HCl, pH 8.0) to 3 ml

ε_M (412 nm) (thionitrobenzene) = 13,600.

2.2.6.4 p-NPGB (Schleuning and Fritz, 1976)

Assay conditions: p-NPGB 10 µl

enzyme 10 - 100 µl

assay buffer (0.1 M sodium barbitol, pH 8.3) 2 ml.

p-NPGB was prepared as a 0.05 M solution in dimethylformamide, then diluted to 0.01 M using acetonitrile.

$$\text{Molarity p-nitrophenol released} = \Delta A_{402} \times 5.46 \times 10^{-5}$$

2.2.7 Inhibitor studies

Factor $\bar{D}$ was incubated with a variety of inhibitors whose activity is directed against different families of proteases. The inhibitors studied and the conditions under which they were incubated with factor $\bar{D}$ are summarised in Table 2.2.

Most of the inhibitors were water soluble, and were prepared in aqueous buffers, usually 0.1 M Tris/HCl/0.2 M NaCl/2 mM EDTA, pH 8.0.

Stock solutions of water-insoluble inhibitors were as follows:

iPr_2P-F: 2.5 M in propan-2-ol

PMSF: 20 mM in propan-2-ol

pNPGB: 20 mM in dimethylformamide

BDC-OH: 0.12 M in acetone

Working solutions were prepared in aqueous buffer. In the case of pNPGB which is hydrolysed spontaneously in nucleophilic buffers, the working solution was prepared in 0.004 M PBS, pH 7.2.

Table 2.2 Protease inhibitors and conditions of incubation with factor $\bar{D}$

Inhibitor	Concentration	Concentration of factor $\bar{D}$	Conditions used
Serine protease inhibitors:			
iPr_2P-F	1 - 10 mM	1.48×10^{-8} M	30 min, 30°C
PMSF	1 mM	4.3×10^{-8} M	60 min, 30°C
p-NPGB	0.5 - 10 mM	"	60 min, 25°C
p-aminobenzamidine	5 mM	"	60 min, 37°C
LBTI	0.5 mg/ml	4.3×10^{-8} M	60 min, 37°C
SBTI (type 1-S)	"	"	"
chick ovomucoid	"	"	"
chick ovoinhibitor	"	"	"
Aprotonin (Trasylol)	"	"	20 min, 30°C
Thiol protease inhibitors:			
BDC-OH	9.52×10^{-5} M	1.0×10^{-7} M	30 min, 25°C
IAA	20 mM	5.7×10^{-8} M	60 min, 25°C
Leupeptin	1 mg/ml	4.3×10^{-8} M	20 min, 30°C
Carboxyl protease inhibitor:			
Pepstatin	1 mg/ml	4.3×10^{-8} M	20 min, 30°C

Continued.....

Table 2.2 Continued

Inhibitor	Concentration	Concentration of factor $\bar{D}$	Conditions used
Plasma protease inhibitors:			
α_2-macroglobulin	2.5×10^{-7} M	4.3×10^{-8} M	60 min, 37°C
inter-α- trypsin inhibitor	"	"	"
$C\bar{1}$ - INH	"	"	"
AT - III	1 mg/ml	"	"
Synthetic substrates:			
p-nitroanilides	0.5 mM	5.0×10^{-8} M	30 min, 37°C
TAME	25 mM	"	60 min, 37°C
ALME	"	"	"
AGLME	"	"	"

The general method used was to incubate equal volumes of factor $\bar{D}$ (in 0.1 M Tris/HCl/0.2 M NaCl/ 2 mM EDTA, pH 8.0) and inhibitor solution. Following incubation under the stated conditions, residual factor $\bar{D}$ haemolytic activity was assayed. Normally, incubations with inhibitors and also haemolytic assays were performed in duplicate. Controls of inhibitor alone, and of inhibitor solvent plus factor $\bar{D}$, where appropriate, were also included.

A number of the inhibitors required special incubation conditions. Incubation with iPr_2P-F or pNPGB was followed by dialysis for 20 h at 4°C against inhibitor-free buffer or else samples were stood at 4°C for 24-48 h to allow hydrolysis of excess inhibitor. Studies with BDC-OH (Rohrbach et al, 1973) were carried out in 0.05 M sodium acetate/acetic acid buffer, pH 5.0 and factor $\bar{D}$ was equilibrated in this buffer before incubation with BDC-OH. This was necessary since the carbonium ion of BDC-OH, which is the active sulphydryl reagent, has maximum stability at pH 5.1 $\pm$ 0.1.

2.2.8 Miscellaneous studies on factor $\bar{D}$ haemolytic activity

2.2.8.1 Trypsin-inducible factor $\bar{D}$ activity

The method described by Fearon et al (1974) was used. Factor $\bar{D}$ was obtained either by ion-exchange chromatography on QAE-Sephadex A-50 (section 2.2.4.1) or from gel filtration of serum on Sephadex G-75 (section 2.1.5.2). Factor $\bar{D}$ was inhibited with 2 mM iPr_2P-F by incubating at 30°C for 30 min. The inhibited samples were stood for 24 h at 4°C and diluted five times in 0.004 M PBS, pH 7.2 before assaying.

To 0.4 ml of the diluted iPr_2P-F-treated factor $\bar{D}$ (0.2 - 2 µg) was added 0.1 ml trypsin (0.05 mg/ml). The mixture was incubated at 37°C for 30 sec. At the end of this time, 0.1 ml SBTI (0.5 mg/ml) was added, and the samples transferred immediately to ice before assaying for haemolytic activity. All solutions were in 0.004 M PBS, pH 7.2 except for factor $\bar{D}$ from column fractions which were in the elution buffer for QAE-Sephadex A-50 (section 2.2.4.1).

2.2.8.2 Collagenase digestion of factor $\bar{D}$

Factor $\bar{D}$ (75 ng) and collagenase (0.001 - 10 ng) in a total volume of 35 µl of 0.01 M Tris/HCl buffer, pH 7.4 containing 0.15 M NaCl and 5 mM $CaCl_2$, were incubated for 20 h at 37°C. The digest was assayed for factor $\bar{D}$ activity.

2.2.9 Automated analysis of amino acids and amino sugars

2.2.9.1 Preparation of samples for analysis

(i) Hydrolysis in 6 M HCl Protein or peptide samples (0.5 - 1.0 nmoles) for hydrolysis were dried *in vacuo* in pyrex tubes. The dried samples were redissolved in 0.25 ml 6 M redistilled HCl containing 4 mM phenol. Dissolved gases were removed by repeated freezing and thawing under vacuum by the method of Moore and Stein (1963). After sealing the tubes under vacuum, samples were hydrolysed usually for 48 h at 110°C. For small peptides, 24 h was considered sufficient for complete hydrolysis. Cystine was estimated as S-carboxymethyl-cysteine after acid hydrolysis of reduced and alkylated samples in the presence of 0.05% (v/v)

2-mercaptoethanol. Samples lacking S-carboxymethyl-cysteine (as detected by liquid scintillation counting of ^{14}C-labelled samples) were not completely evacuated but were flushed with nitrogen before sealing. No corrections were made for the destruction of serine or threonine.

(ii) Preservation of Tryptophan Protein samples (2-3 nmoles) were hydrolysed in 3 M mercaptoethane-sulphonic acid (50 µl) for 24 h at 110°C in evacuated and sealed tubes (Penke et al, 1974). After 24 h, 4 M NaOH (25 µl) was added to the samples. Control samples of 3 nmoles and 30 nmoles lysozyme were hydrolysed simultaneously and gave an approximately 75% yield of tryptophan. Analyses of factor $\bar{D}$ were corrected assuming a similar yield. Lysozyme was considered a suitable control as both this enzyme and probably factor $\bar{D}$ (section 3.5) contain less than 1.5% (w/w) carbohydrate.

(iii) Preservation of amino-sugars Samples (2-3 nmoles) were incubated in 3 M p-toluene sulphonic acid (50 µl) for 24 h at 110°C in sealed, evacuated tubes (Allen and Neuberger, 1975). At the end of this period, the hydrolysates were neutralised with 4 M NaOH (25 µl). Control samples of lysozyme (3 nmoles) with added glucosamine and galactosamine (10 nmoles of each) were hydrolysed concurrently in the same way. The yields of amino sugars compared with the theoretical values were 50% (glucosamine) and 30% (galactosamine). The values obtained for factor $\bar{D}$ were corrected accordingly.

2.2.9.2 Automated analysis of hydrolysates

Hydrolysed samples were analysed on a Durrum D-500 amino acid analyser. Acidic and neutral amino acids were eluted with sodium citrate buffers (0.2 M with respect to sodium) at pH 3.27 and pH 4.25, and basic amino acids were eluted with sodium citrate buffer (1.1 M with respect to sodium) at pH 7.90.

The following buffer programmes were used:-

(i) Amino acids (except for hydroxylysine and hydroxyproline)

Sample injection:	0 min
pH 3.27 buffer:	0 - 27.0 min
pH 4.25 buffer:	27.0 - 45.5 min
pH 7.90 buffer:	45.5 - 82.0 min
Temperature change (from 50°C to 65°C):	33.0 min

This programme was also used for tryptophan analysis. Tryptophan elutes from the column in the pH 7.90 buffer immediately after the ammonia peak (64.7 min after sample injection).

(ii) Hydroxyproline and hydroxylysine

Sample injection	0 min
pH 2.85 buffer:	0 - 15.0 min
pH 3.27 buffer:	15.0 - 35.0 min
pH 4.25 buffer:	35.0 - 53.5 min
pH 7.90 buffer:	53.5 - 87.5 min
Temperature change (from 50°C to 65°C):	15.0 min

Hydroxyproline elutes in the pH 2.85 buffer ahead of both methionine sulphone or aspartic acid (19.4 min after sample injection). Hydroxylysine elutes between histidine and lysine in the pH 7.90 buffer (65.5 min after sample injection).

(iii) Amino sugars

Sample injection:	0 min
pH 3.27 buffer:	0 - 12.0 min
pH 4.25 buffer:	12.0 - 32.5 min
pH 7.90 buffer:	32.5 - 72.0 min
Temperature change (from 50°C to 65°C):	30.0 min

Glucosamine and galactosamine elute from the column just ahead of histidine, at 39.8 min and 40.8 min respectively after sample injection.

2.2.10 Chemical modification of factor $\bar{D}$

2.2.10.1 Reduction and alkylation

In the method which follows, the quantities stated are for use with 100 nmoles of protein. In practice, preparations were used containing 70-150 nmoles of factor $\bar{D}$. To decrease the loss of protein through binding to glass, all glassware was siliconised before use with 2% (v/v) surfasil in petroleum ether.

Factor $\bar{D}$ (100 nmoles) was dissolved in 6 M guanidine/HCl/0.4 M Tris/HCl/2 mM EDTA, pH 8.0 (1.5 ml) containing 20 mM dithiothreitol, and incubated in a sealed flask at 37°C for 3 h with frequent mixing. In preparations where

cysteine was radioactively labelled, iodo-[2-^{14}C]-acetic acid (100 µCi in 100 µl distilled water) was added and the incubation was continued for 5 min at 25°C. Ice-cold carrier iodoacetic acid was added to a concentration of 60 mM, and the mixture was cooled on ice for 45 min. The reduced and alkylated sample was then dialysed at 4°C against distilled water (1 litre) for 3 h followed by two changes of 0.5 M acetic acid (1 litre) and was freeze-dried.

2.2.10.2 Labelling of factor $\bar{D}$ with [1,3-^{14}C]-iPr_2P-F

Factor $\bar{D}$ (50 nmoles) in 5 ml, 0.01 M Tris/HCl/0.15 M NaCl/ 1 mM $CaCl_2$/ 1 mM $MgCl_2$, pH 8.0, was incubated at 30°C for 1 h with 1 µmol of [1,3-^{14}C]-iPr_2P-F (100 µCi). Non-radioactive iPr_2P-F was added to give a final concentration of 10 mM and the mixture was incubated for a further 1 h at 30°C. The inactivated factor $\bar{D}$ was dialysed for 24 h against three changes of 0.5 M acetic acid and was freeze-dried.

2.2.10.3 Succinylation of amino groups

Factor $\bar{D}$ (100 nmoles) was dissolved in 2 ml 1 M Tris/HCl buffer, pH 10.5, and gently mixed for 30 min at 25°C to ensure maximal solution of the freeze-dried protein. Succinic anhydride was then added, with continued mixing, in aliquots of 10 mg at 10 min intervals until a total of 50 mg had been added. This gave a final molar excess of succinic anhydride over factor $\bar{D}$-lysine of 600-fold. Stirring was continued for a further 60 min. The

succinylated sample was then dialysed for 20 h at 4°C against two changes of 5 mM NH_3, and was freeze-dried.

2.2.11 Protein fragmentation and purification of peptides

2.2.11.1 Cleavage at methionyl bonds using cyanogen bromide

The method of Gross (1967) was followed. A 500-fold molar excess of cyanogen bromide over factor $\bar{D}$-methionine was used. Reduced and alkylated factor $\bar{D}$ (40-130 nmoles) was dissolved in 70% (v/v) formic acid (0.7 ml/100 nmoles factor $\bar{D}$) containing 0.215 M CNBr, and incubated for 16-20 h in the dark at 25°C. The mixture was freeze-dried and then redissolved in 70% (v/v) formic acid (0.7 ml) and applied to a column (1.5 cm x 100 cm) of Sephadex G-50 superfine equilibrated with 5% (v/v) formic acid. The flow rate was 2.5 ml/h. Peptides eluting near the void volume on the Sephadex G-50 column were further purified by gel filtration on a column (1.5 cm x 100 cm) of Sephadex G-75 superfine equilibrated with 10% (v/v) formic acid and run at 1.5 ml/h. Smaller peptides were further purified by high pressure liquid chromatography (HPLC) using a μBondapak C-18 column. Two solutions were used in the HPLC of peptides: solution A, which was composed of CH_3CN and 10 mM NH_4HCO_3 (5:95, v/v) and solution B, which was composed of CH_3CN and 10 mM NH_4HCO_3 (64:40, v/v). The column was equilibrated with a mixture of solution A and solution B in the ratio 95:5 (v/v) and then eluted with a linear gradient to give a final ratio of 10:90 (v/v) of solution A to solution B. The column was pumped at 1.5 ml/h by a Waters Associates system (model 6000A).

2.2.11.2 Cleavage at tryptophanyl bonds using o-iodoso-benzoic acid

(i) Pilot study O-iodosobenzoic acid (Mahoney and Hermodson, 1979) was used in a pilot study to test its suitability as a reagent for cleaving factor $\bar{D}$ at tryptophan residues.

Factor $\bar{D}$ (0.2 mg) and o-iodosobenzoic acid (0.4 mg) in 0.6 ml of a solution of 38.2% (w/v) guanidine hydrochloride in 80% (v/v) glacial acetic acid, were incubated in a sealed flask in the dark for 24 h at 25°C. The digested sample was dialysed for 20 h at 4°C against two changes of 0.5 M acetic acid (1 litre) and was freeze-dried. After redissolving in 0.1 M Tris/HCl/4 M urea/1% (w/v) SDS, pH 8.0 (200 µl), a sample (20 µg) was taken for SDS/20% (w/v) polyacrylamide slab-gel electrophoresis and the remainder of the sample (180 µg) was electrophoresed on SDS/12% (w/v) polyacrylamide cylindrical gels. After staining with Coomassie Blue, the cylindrical gel was sliced and the radioactivity counted (section 2.2.1.6).

(ii) Large-scale digest Two large-scale digests were carried out using (1) 70 nmoles and (2) 26 nmoles of factor $\bar{D}$. The factor $\bar{D}$ was digested with o-iodosobenzoic acid as described for the pilot study. Following digestion for 24 h the samples were not dialysed, but were applied directly to a column (1.5 cm x 100 cm) of Sephadex G-50 superfine equilibrated with 5% (v/v) formic acid. The peptides were eluted at 5 ml/h. Larger peptides from the 70 nmole digest were repurified in Sephadex G-75 superfine equilibrated with 10% (v/v) formic acid and run at 3 ml/h.

The salt peaks from both digests were pooled and further purified and desalted by gel filtration on a column (0.9 cm x 60 cm) of Sephadex G-10 equilibrated with 0.1 M NH_4HCO_3 and run at a flow rate of 5 ml/h.

2.2.11.3 Cleavage at arginyl bonds using trypsin-TPCK

Succinylated factor $\bar{D}$ (90 nmoles) labelled with iodo-[2-^{14}C]-acetic acid and trypsin-TPCK (40 µg in 1 mM HCl) in 1.45 ml 0.1 M NH_4HCO_3 were incubated with occasional mixing for 4 h at 37°C. At the end of this time, 2.5 mM iPr_2P-F (20 µl) was added to give a 30-fold excess of iPr_2P-F over trypsin-TPCK. The mixture was centrifuged at 45,000 g for 20 min in a Beckman L-2 ultracentrifuge. The pellet was washed in a further 0.5 ml of 0.1 M NH_4HCO_3, recentrifuged, and the combined supernatants were applied to a column (1.5 cm x 100 cm) of Sephadex G-75 superfine equilibrated with 0.1 M NH_4HCO_3. The major peptides were repurified by HPLC using a µBondapak C-18 column (0.39 cm x 30 cm) equilibrated with 0.1% (w/v) NH_4HCO_3 containing 5% (v/v) CH_3CN. The peptides were eluted over a period of 1 h in a linear gradient from 5% to 65% (v/v) CH_3CN in 0.1% (w/v) NH_4HCO_3. The column was pumped at 1.0 ml/min by a Waters Associates system (model 6000A).

2.2.11.4 Cleavage at glutamic acid residues using V8 protease

The peptide, CNBr-1, obtained by cyanogen bromide cleavage of factor $\bar{D}$ (section 2.2.11.1) was digested further with protease from S. aureus V8 (Drapeau et al, 1972). The peptide (100 nmoles) and V8 protease (0.08 mg) in 0.1 M $NaHCO_3$ (1.2 ml) were incubated with occasional mixing for 8 h at 37°C. The resultant peptide mixture was freeze-dried and then purified directly by HPLC using the system described in section 2.2.11.3, except that the gradient was run from 5% to 55% (v/v) CH_3CN in 0.1% (w/v) NH_4HCO_3.

2.2.11.5 C-terminal analysis using carboxypeptidase-Y

The method of Hayashi (1977) was used. Protein or peptide (5 nmoles) and carboxypeptidase-Y (5-10 µg) in 0.2 M acetic acid adjusted to pH 5.5 with pyridine (100 µl) were incubated at timed intervals from 0-5 h at 25°C. Following incubation for the appropriate time, samples were boiled for 5 min to inactivate carboxypeptidase-Y and freeze-dried before amino acid analysis.

2.2.12 Amino acid sequence determination

2.2.12.1 N-terminal amino acid analysis

Normally, N-terminal residues were determined by automated sequencing, except in the case of peptides which could be identified from their amino acid composition. In these cases, the dansyl chloride method of N-terminal analysis (Hartley, 1970) was used.

The freeze-dried peptide (1-5 nmoles) was dissolved in 0.2 M $NaHCO_3$ (10 µl) and dried *in vacuo*. The peptide was dansylated by incubation with 10 µl distilled water and 10 µl dansyl chloride (2.5 mg/ml in acetone) at 37°C for 1 h. The sample was again dried *in vacuo* and then hydrolysed in a sealed tube at 110°C for 18 h in 6 M HCl (50 µl). The hydrolysed sample was dried *in vacuo*.

Dansyl amino acids were identified by chromatography on polyamide sheets (5 cm x 5 cm) (Woods and Wang, 1967). The samples were redissolved in 50% (v/v) pyridine (10 µl) and applied to one corner of the sheet, with equal amounts on both sides, and the origins superimposed. To the origin on one side, 1 µl of a standard dansyl amino acid mixture (1 mg/ml of each of dansyl-phenylalanine,-isoleucine, -proline,-glutamic acid,-glycine,-serine and -arginine) was applied and the plates developed by ascending chromatography using the following solvent systems:-

(1) 1.5% (v/v) formic acid

(2) toluene : acetic acid, 4:1 (v/v)

(3) ethyl acetate : acetic acid : methanol, 20:1:1 (v/v)

(4) 0.05 M trisodium phosphate in 25% (v/v) aqueous ethanol.

Solvents 2-4 were run at right angles to solvent 1, and solvent 4 was used only when chromatography in the first three solvents did not give conclusive results.

The dansyl amino acids were observed under an ultraviolet lamp (350 nm) after running in each solvent, and the residues were identified by comparison with standard dansyl amino acids.

2.2.12.2 Sequence analysis by automated Edman degradation

Automated Edman degradation (Edman and Begg, 1967) was carried out in a Beckman 890C sequenator using the 0.3 M Quadrol programme of Hunkapiller and Hood (1978). Polybrene (2 mg) was added to the sequencer cup prior to the application of each sample (Klapper et al, 1978).

The thiazolinones released in the sequencer were converted into the amino acid phenylthiohydantoin derivatives by heating at 80°C under nitrogen for 10 min in 1 M HCl (200 µl) containing 0.1% (v/v) ethanethiol. After two extractions with ethyl acetate (700 µl), 20-80% of the extracted amino acid phenylthiohydantoin derivative was examined by high pressure liquid chromatography. The chromatography was done on a µBondapak C-18 column using a 14-56% methanol gradient in aqueous buffer (Bridgen et al, 1976) pumped at 2.2 ml/min by a Waters Associates system (model 6000A) with a UV detector. The acid layer from the ethyl acetate extraction was dried and examined in the HPLC system described above, using a 14-30% methanol gradient. The presence of ^{14}C-labelled S-carboxy-methyl-cysteine or iPr_2P-serine was confirmed by liquid scintillation counting.

The stepwise yield for a given PTH amino acid derivative was calculated using the formula

$$R_y = \left[\frac{Y_B}{Y_A} \right]^{\frac{1}{b'-a'}}$$

where the sequence is xxxxx a xxxx b xxx. Y_A and Y_B are the yields for residues a and b respectively, and a' and b' are the cycle numbers of the residues.

CHAPTER 3

PURIFICATION AND CHARACTERISATION OF FACTOR $\bar{D}$

3.1 INTRODUCTION

The study of factor $\bar{D}$ has been hampered by its low plasma concentration (1-2 μg/ml) and by the difficulty in removing contaminants of similar molecular weight and behaviour on ion-exchange columns. Some of these contaminating proteins appear to be generated during the purification procedure or during storage of plasma or partially purified factor $\bar{D}$ (section 3.2; Davis *et al*, 1979b). Several methods of purifying factor $\bar{D}$ have been published but only more recently have purification procedures been developed for obtaining highly purified factor $\bar{D}$ (Lesavre *et al*, 1979; Davis *et al*, 1979b; Volanakis *et al*, 1980). Only one of these methods has been used for extensive sequence studies (Volanakis *et al*, 1980) and this method required a preparative gel electrophoresis step in order to obtain sufficiently pure factor $\bar{D}$ to give unambiguous sequence data.

At the time of embarking on the work described here, no suitable purification method was available for use in sequence studies. Therefore, the aim of this section of the work was to develop such a method and to confirm and extend earlier miscellaneous observations on the nature of factor $\bar{D}$.

3.2 PURIFICATION OF FACTOR $\bar{D}$

3.2.1 Preliminary studies

A variety of approaches was tried in order to obtain pure factor $\bar{D}$. The initial approach was based on the method of Hunsicker et al (1973). This involved fractionation of a pH 5.4 euglobulin precipitate on DEAE-cellulose, followed by gel filtration on Sephadex G-200.

(1) Euglobulin precipitation

Euglobulin precipitation is a convenient preliminary step in the purification of complement, most early components remaining in the pseudoglobulin fraction, while C1 and properdin are precipitated as euglobulin under suitable conditions of pH and ionic strength (Lepow et al, 1965). Early papers on factor $\bar{D}$ had described it both as a euglobulin (Hunsicker et al, 1973) and as a pseudoglobulin (Dierich et al, 1974).

Euglobulin precipitates were prepared over a pH range from 4.0 to 8.0 (section 2.2.4.2). After dialysis, the precipitates were resuspended, and the haemolytic activity of the resuspended precipitates and the supernatants was measured. No precipitate contained more than 12% of the total factor $\bar{D}$ haemolytic activity. Therefore, the pH 5.4 pseudoglobulin supernatant was used for further purification of factor $\bar{D}$ on CM-Sephadex C-50.

(2) CM-Sephadex C-50

R. Prohaska (unpublished results) observed that factor $\bar{D}$ was eluted in high salt from CM-Sephadex C-50 equilibrated with sodium phosphate buffer, pH 6.0. Since when euglobulin

was fractionated on DEAE-cellulose factor $\bar{D}$ eluted in three separate peaks of haemolytic activity (Hunsicker et al, 1973), it seemed that a CM-Sephadex C-50 step in which factor $\bar{D}$ haemolytic activity eluted as one broad peak, might be more suitable for the initial fractionation of the pseudoglobulin. It also had the advantage that C2 and factor B could be sequentially recovered from the same column (Kerr and Porter, 1978).

(3) $(NH_4)_2SO_4$ fractionation

When whole serum was fractionated by $(NH_4)_2SO_4$ precipitation, half of the factor $\bar{D}$ precipitated in the 30-50% saturation fraction, the remainder being in the supernatant from the 50% saturation fraction. However, using partially purified factor $\bar{D}$, 80% of the haemolytic activity was recovered in the 50-70% saturation fraction. The reason for this discrepancy was not apparent, but could be explained if factor $\bar{D}$ associated with another protein in whole serum, as suggested by Konno et al (1978). Although the precipitation step was not essential, it was a convenient way of concentrating the large volume of the factor $\bar{D}$ pool obtained from CM-Sephadex C-50, and was therefore included in the method which was finally adopted.

(4) Removal of low molecular weight contaminants

The crude factor $\bar{D}$ obtained from CM-Sephadex C-50 was fractionated on a column (5 cm x 100 cm) of Sephadex G-75 and the fractions containing factor $\bar{D}$ haemolytic activity were pooled. Removal of contaminants was initially attempted by repeating the gel filtration step on a smaller

column (2.5 cm x 100 cm) of Sephadex G-75. However SDS/ polyacrylamide gels of these preparations showed that they always contained up to 30% contaminating proteins. In view of later observations (section 3.2.2) and those of Davis et al (1979b) it is possible that some of these contaminants were generated by proteolysis of larger proteins during the purification procedure. Attempts to remove these contaminants included fractionation on DEAE-cellulose DE32, agarose-lysine, CM-cellulose 32 and concanavalin A-Sepharose. Initial amino acid analyses and amino acid sequencer studies were inconclusive and non-reproducible, but success was finally achieved by using CM-cellulose 32 and concanavalin A-Sepharose as described below. The increase in specific activity following fractionation of factor $\bar{D}$ on columns of either DEAE-cellulose 32 or agarose-lysine was not sufficiently great to warrant their inclusion in the purification scheme which was finally adopted.

3.2.2 Routine purification of factor $\bar{D}$

In the procedure described both fresh frozen plasma and outdated plasma or serum gave preparations of factor $\bar{D}$ of comparable yield and specific activity. However, it was preferable to use fresh or fresh frozen plasma since low molecular weight contaminants having similar molecular weight and ion-exchange properties as factor $\bar{D}$ appeared to be generated during storage of plasma or partially purified factor $\bar{D}$ fractions. Addition of protease inhibitors in the early stages of purification did not increase the recovery of factor $\bar{D}$. The results described here refer to preparations using serum prepared from outdated plasma, and are summarised in Table 3.1.

Table 3.1 Purification of factor $\bar{D}$ from serum

Fraction	Total protein (A_{280})	Total haemolytic activity $x10^{-4}$ (units)	Specific activity (units/A_{280})	Purification factor	Yield
Serum (4000 ml)	221 000	38	1.73	1	100
Pseudoglobulin	209 000	35	1.69	0.98	92
CM-Sephadex C-50	422	32	758	440	84
50-70% $(NH_4)_2SO_4$ precipitate	144	26	1790	1040	68
Sephadex G-75	17	17.3	10 200	5890	45
CM-cellulose-32	3.2	9.2	29 600	17 100	24
Concanavalin-A-Sepharose	1.8*	8.6	49 200	28 500	22

*The final yield of protein as estimated by amino acid analysis was 0.9 mg. This gives a final specific activity (units/mg of protein) for factor $\bar{D}$ of 103 000 and a purification factor of approx. 60 000.

After precipitation of euglobulin, more than 99% of the protein remaining in the pseudoglobulin fraction was removed by ion-exchange chromatography on CM-Sephadex C-50. On washing the column with 0.4 M sodium phosphate buffer, pH 6.0, the factor D haemolytic activity eluted above 16 mS in two peaks (Fig. 3.1). The first peak of activity was probably due to enhancement of haemolysis by traces of factor B which had been left on the column after the 0.2 M sodium phosphate wash and was eluted immediately prior to the factor $\bar{D}$ with the 0.4 M buffer. The activity peak was broad and no improvement in purification resulted from eluting factor $\bar{D}$ with a linear gradient instead of the stepwise wash. It was found convenient to concentrate the material from the CM-Sephadex C-50 column by $(NH_4)_2SO_4$ fractionation. Over 80% of the factor $\bar{D}$ haemolytic activity precipitated in the 50-70% saturation fraction.

The $(NH_4)_2SO_4$ precipitate was further purified by gel filtration on Sephadex G-75, where the factor $\bar{D}$ haemolytic activity was separated from a large number of proteins all having apparent molecular weights of >30,000 (Fig. 3.2). Because of the small amount of protein (15-20 mg) in the factor $\bar{D}$ pool from the Sephadex G-75 column it was found convenient to combine the G-75 factor $\bar{D}$ pools from two preparations before carrying out the final stages of purification. The partially purified factor $\bar{D}$ from Sephadex G-75 was therefore stored at $-70^{o}C$ until use.

Following gel filtration on Sephadex G-75, the preparation appeared to be functionally pure with respect to factor $\bar{D}$

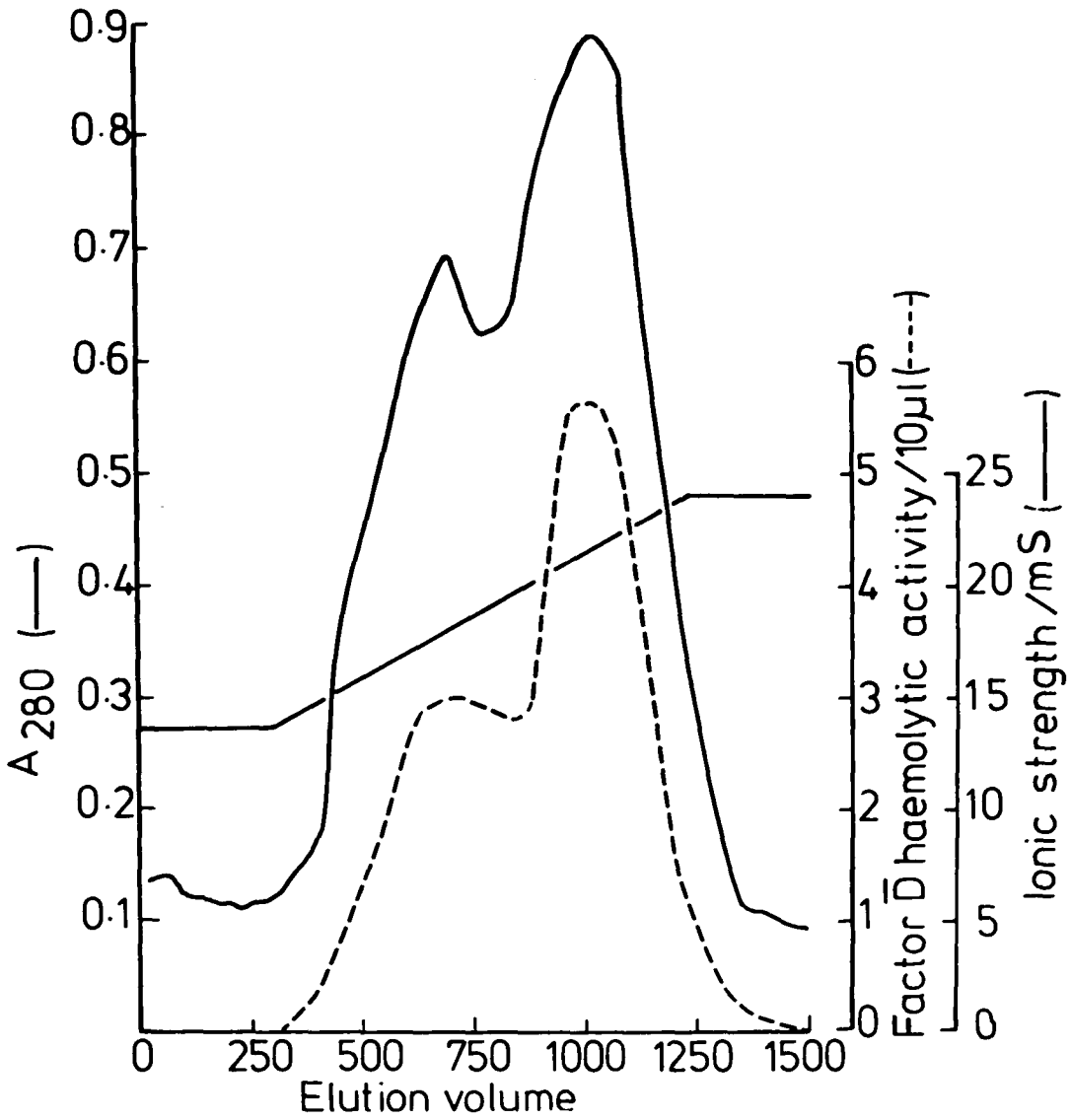


Fig. 3.1 Ion-exchange chromatography of factor $\bar{D}$ on CM-Sephadex C-50

4000 ml of the supernatant (approx. 208,000 A_{280} units) obtained after euglobulin precipitation, was adjusted to 12-14 mS with 1 M sodium phosphate buffer, pH 6.0 and then applied to a column (18 cm x 8 cm) of CM-Sephadex C-50 equilibrated in 0.2 M sodium phosphate buffer, pH 6.0. The column was developed as described in the text. Factor $\bar{D}$ eluted above 16 mS.

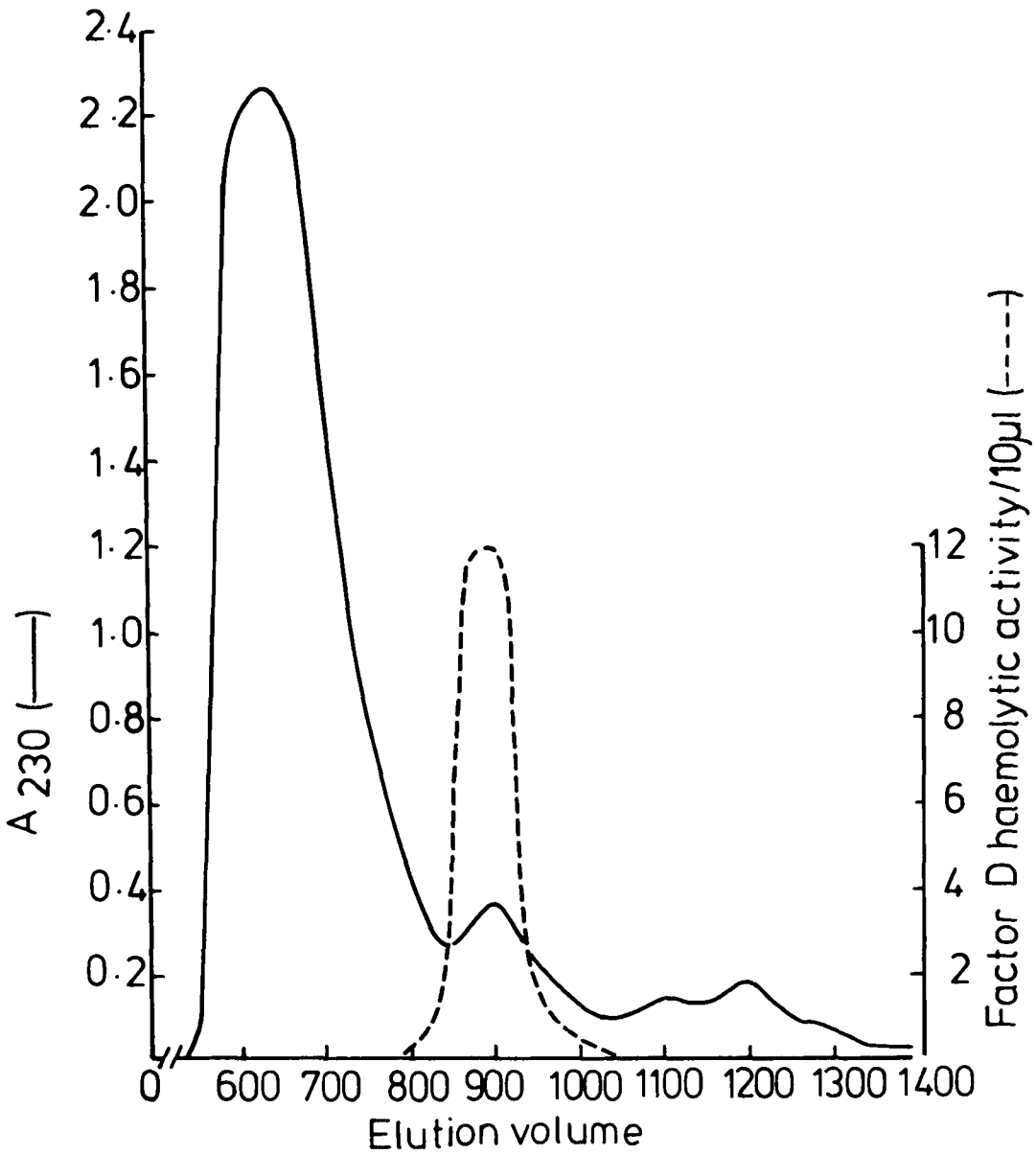


Fig. 3.2 Gel filtration of factor $\bar{D}$ on Sephadex G-75

The precipitate (approx. 146 mg protein) obtained after $(NH_4)_2SO_4$ precipitation was dissolved in 40 ml of 0.1 M Tris/HCl/0.2 M NaCl/2 mM EDTA, pH 8.0 and, after centrifugation, applied to a column (5 cm x 10 cm) of Sephadex G-75 equilibrated with the same buffer.

haemolytic activity, but on further purification on CM-cellulose 32 at least four protein peaks were eluted (Fig. 3.3 (a) and (b)). The factor $\bar{D}$ haemolytic activity eluted between 22-27 mS. Another protein, which was removed by use of Concanavalin A-Sepharose, was eluted between 25-28 mS in the gradient on CM-cellulose 32. This protein could not be distinguished from factor $\bar{D}$ in non-reducing conditions on SDS/polyacrylamide gels, but in reducing conditions had a higher apparent molecular weight compared with factor $\bar{D}$ (Fig. 3.3 (b)). No factor $\bar{D}$ haemolytic activity was associated with this contaminating protein, the concentration of which varied considerably from preparation to preparation, sometimes comprising 50% of the protein in the partially purified factor $\bar{D}$ pool from CM-cellulose 32. Thus the CM-cellulose 32 pool was routinely re-purified on a column of Concanavalin A-Sepharose (Fig. 3.4). Factor $\bar{D}$ was not retained on this column but the contaminant was bound and could only be eluted with 1-O-methyl-α-D-glucopyranoside.

The purified factor $\bar{D}$ appeared to be homogeneous on SDS/polyacrylamide gels in reducing and non-reducing conditions (Fig. 3.5 (a)). On staining the gels with Coomassie Blue, the protein band appeared broad in the non-reduced sample, but amino acid sequence analysis has shown the presence of only one sequence and the numbers of peptides obtained after chemical or enzymatic fragmentation of factor $\bar{D}$ are also consistent with there being only one type of polypeptide chain in the final preparation. A single peak of haemolytic activity which coincided with the single Coomassie Blue-staining band was found following elution of

Fig. 3.3

(a) Ion exchange chromatography of factor $\bar{D}$ on CM-cellulose 32

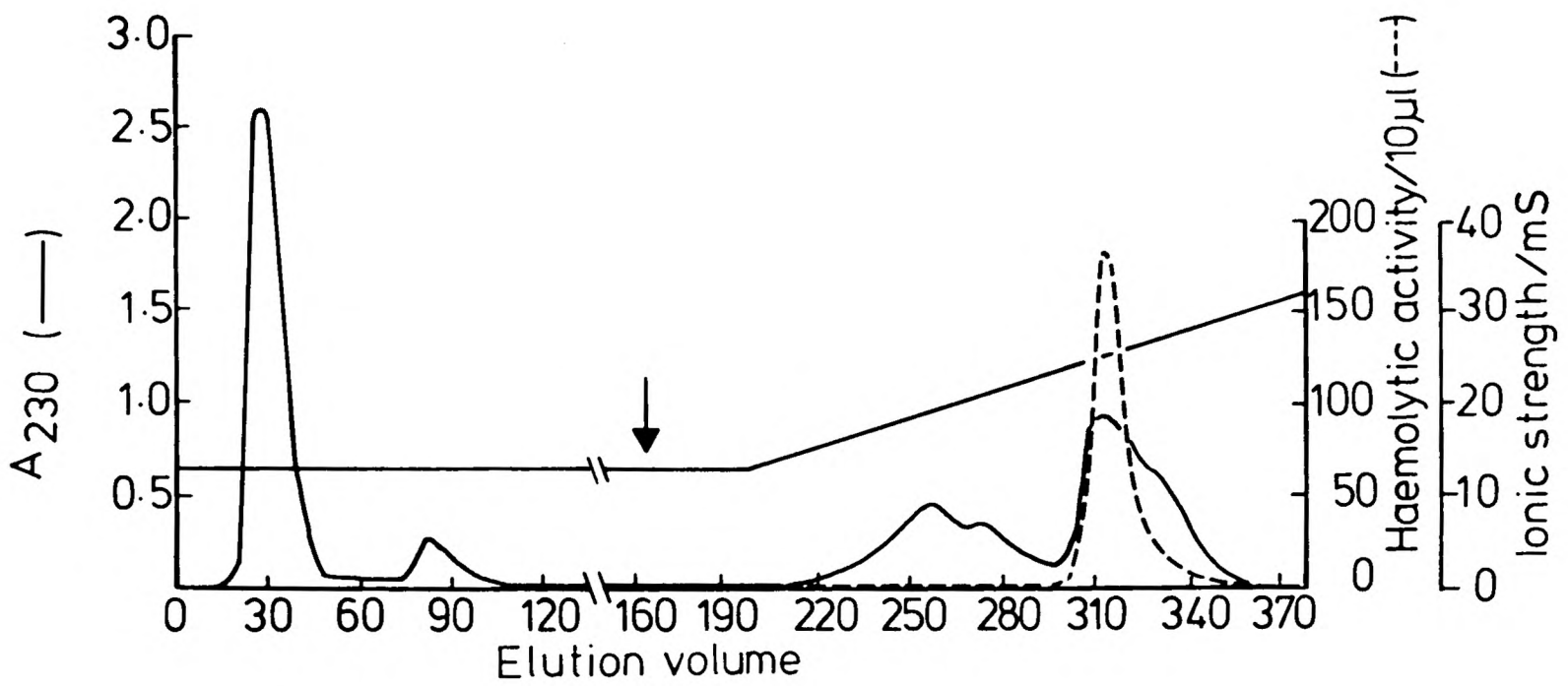
The partially purified factor $\bar{D}$ (approx. 30 mg protein)* from Sephadex G-75 was dialysed against 0.23 M acetic acid/NaOH buffer, pH 5.2, and loaded on to a column (1.5 cm x 20 cm)of CM-cellulose 32 equilibrated in the same buffer. The column was developed as described in the text. Factor $\bar{D}$ haemolytic activity eluted between 22-27 mS. The arrow marks the start of the salt gradient.

*This represents the material from two separate factor D preparations.

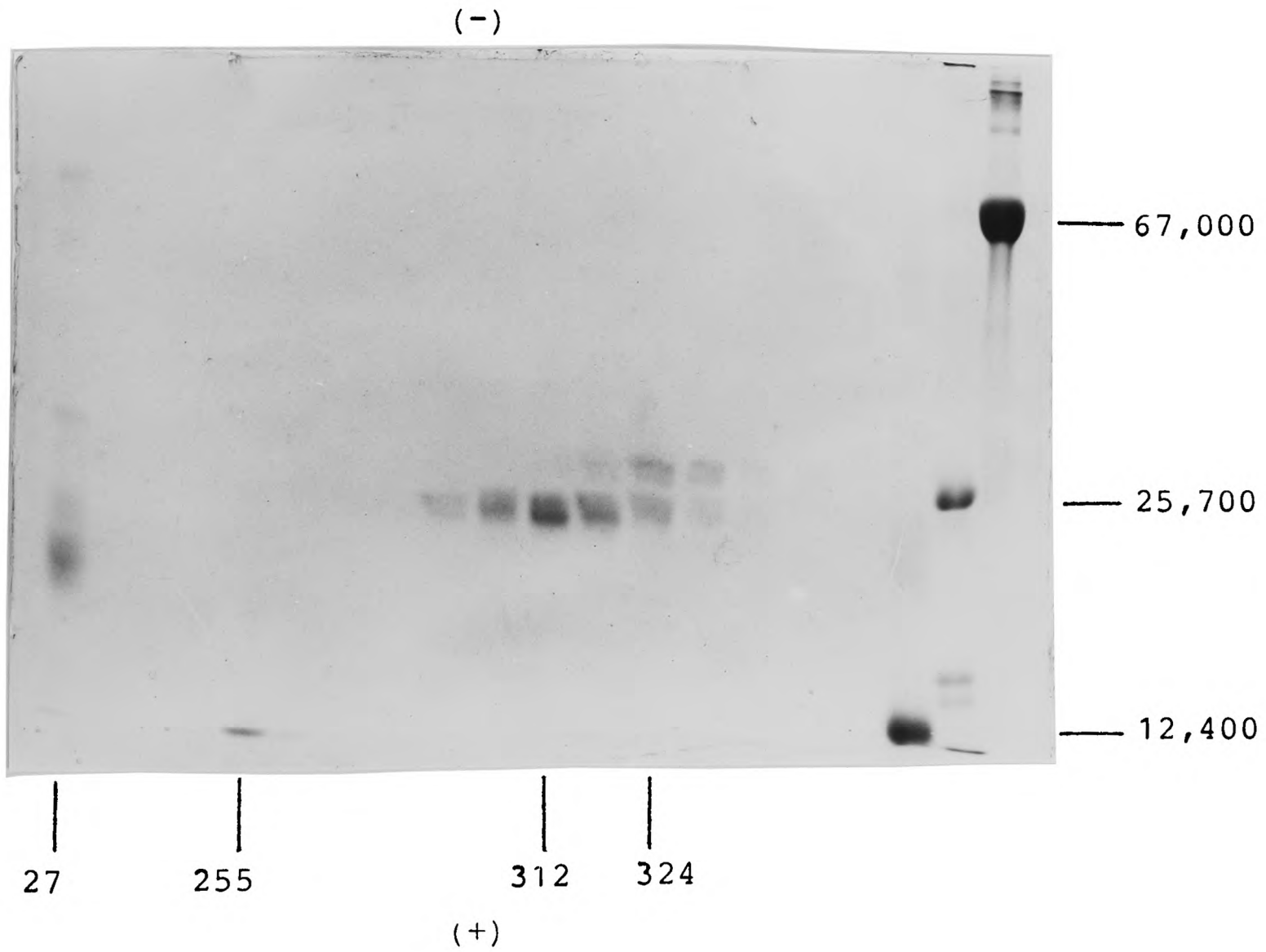
(b) SDS/polyacrylamide gel electrophoresis of column fractions

Samples (25 µl) were taken from column fractions corresponding to elution at 27, 81, 210, 240, 255, 270, 285, 294, 300, 306, 312, 318, 324, 330, 345 and 360 ml on the chromatogram profile shown in (a). The samples were reduced and alkylated and run on a SDS/15% (w/v) polyacrylamide slab gel (section 2.2.2.1). The major contaminants eluted at peak volumes of 27, 255 and 324 ml and factor $\bar{D}$ eluted at a peak volume of 312 ml. Albumin (mol. wt. 67,000), chymotrypsinogen (mol. wt. 25,700) and cytochrome c (mol. wt. 12,400) were reduced and alkylated and run as marker proteins.

(a)


3·0
2·5
2·0
1·5
1·0
0·5
A230 (——)
0 30 60 90 120 160 190 220 250 280 310 340 370
Elution volume
200
150
100
50
0
Haemolytic activity/10µl (---)
40
30
20
10
0
Ionic strength/mS

(b)


Mol.wt.
(-)
67,000
25,700
12,400
27
255
312
324
(+)

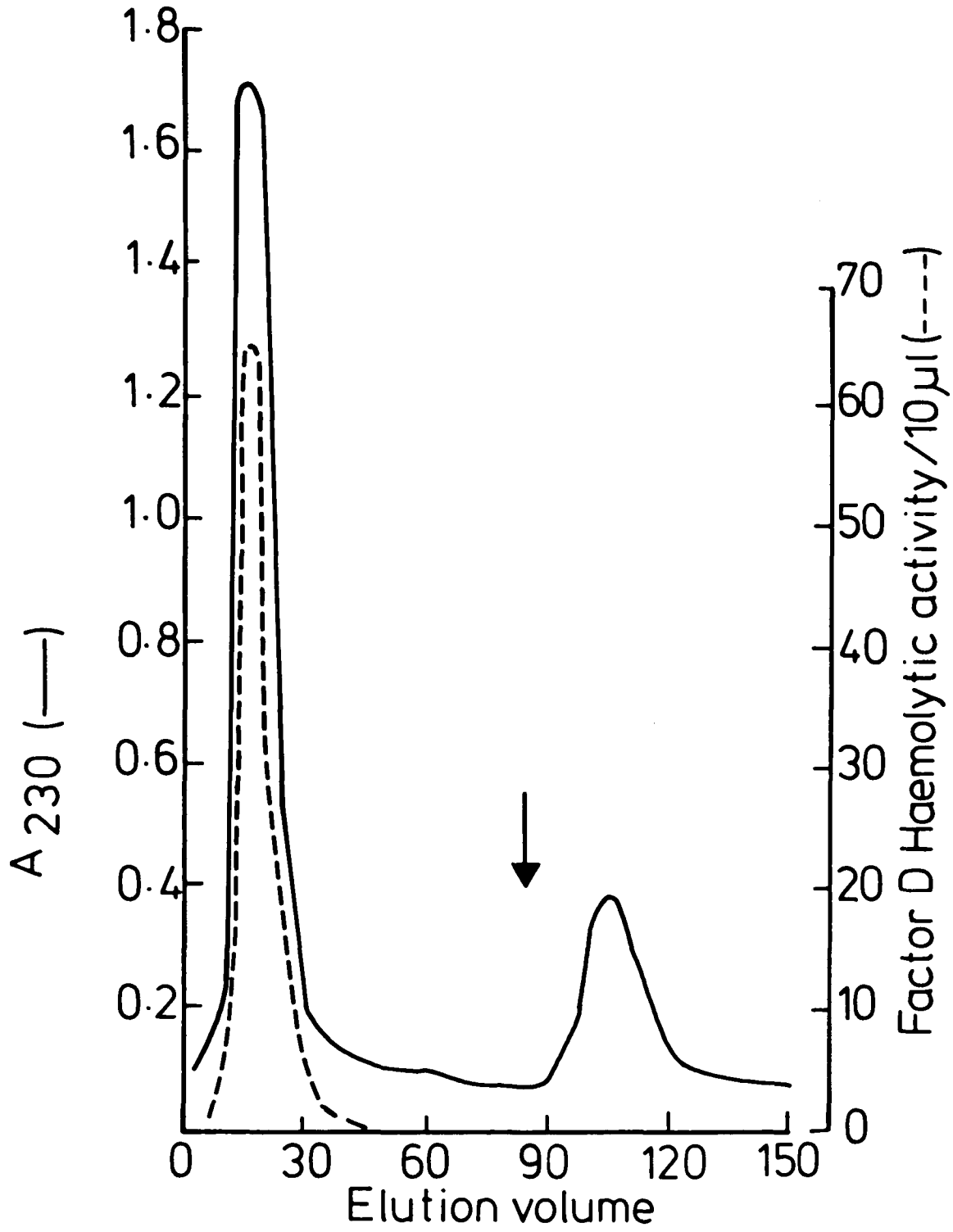


Fig. 3.4 Purification of factor $\bar{D}$ on Concanavalin A-Sepharose

Partially purified factor $\bar{D}$ from the CM-cellulose 32 column (approx. 3.4 mg protein)* in 10 ml of 0.01 M Tris/HCl/0.15 M NaCl/1 mM $MgCl_2$/1 mM $CaCl_2$, pH 8.0, was applied to a column (1 cm x 12 cm) of Concanavalin A-Sepharose equilibrated with the same buffer. The arrow marks the point at which elution with 2.5% (w/v) 1-O-methyl-α-D-glucopyranoside was begun. Factor $\bar{D}$ elutes in the starting buffer.

*This represents the material from two separate factor $\bar{D}$ preparations.

Fig. 3.5 SDS/polyacrylamide gel electrophoresis of purified factor $\bar{D}$

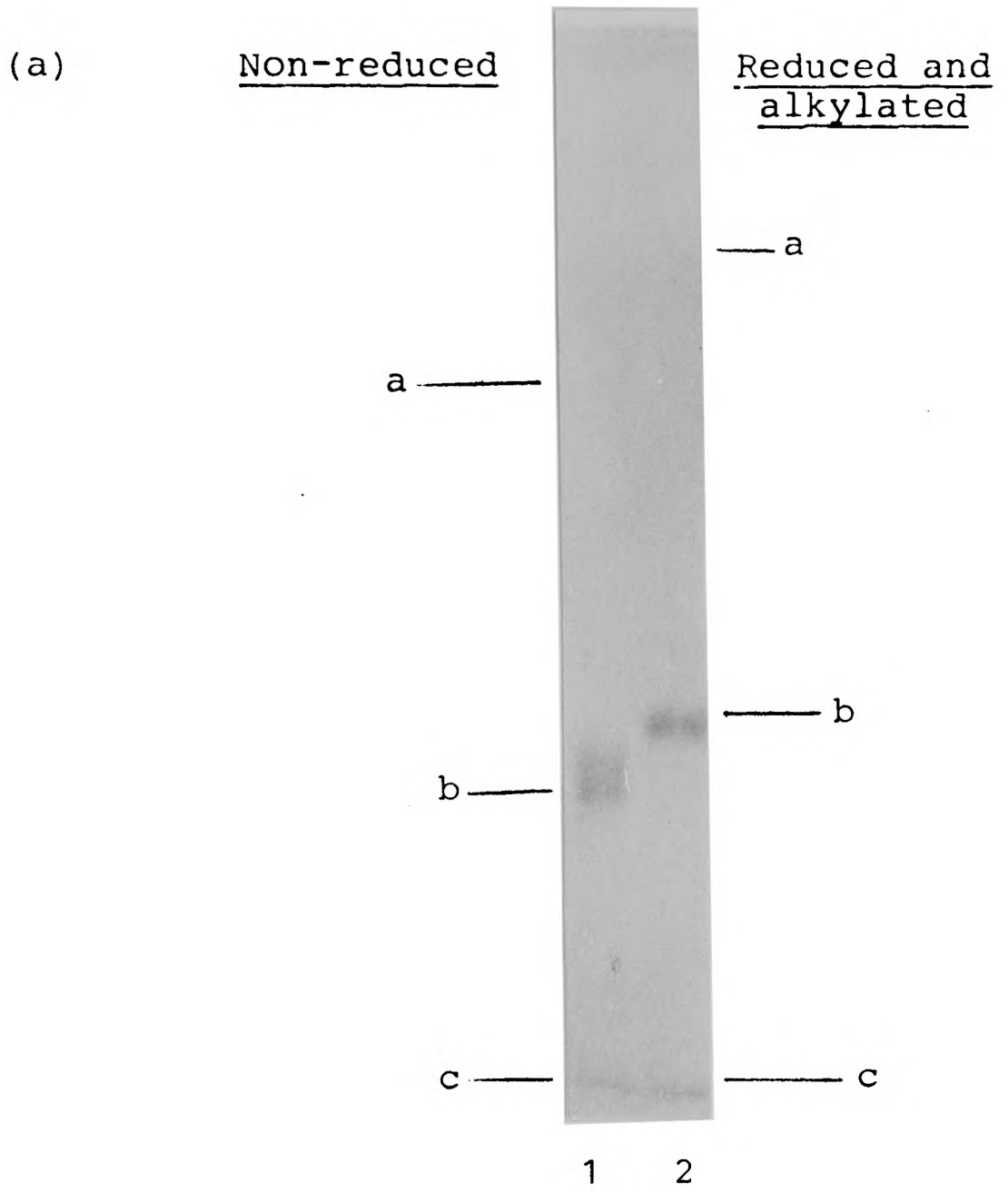
(a) Coomassie blue staining

Factor $\bar{D}$ (5-10 µg), after the final purification step on Concanavalin A-Sepharose was run on a SDS/15% (w/v) polyacrylamide slab gel (section 2.2.2.1). Both nonreduced (track 1) and reduced (track 2) samples were run. Positions of nonreduced and reduced markers are also shown: (a) albumin (mol. wt. 67,000), (b) chymotrypsinogen (mol. wt. 25,700), (c) cytochrome c (mol. wt. 12,400).

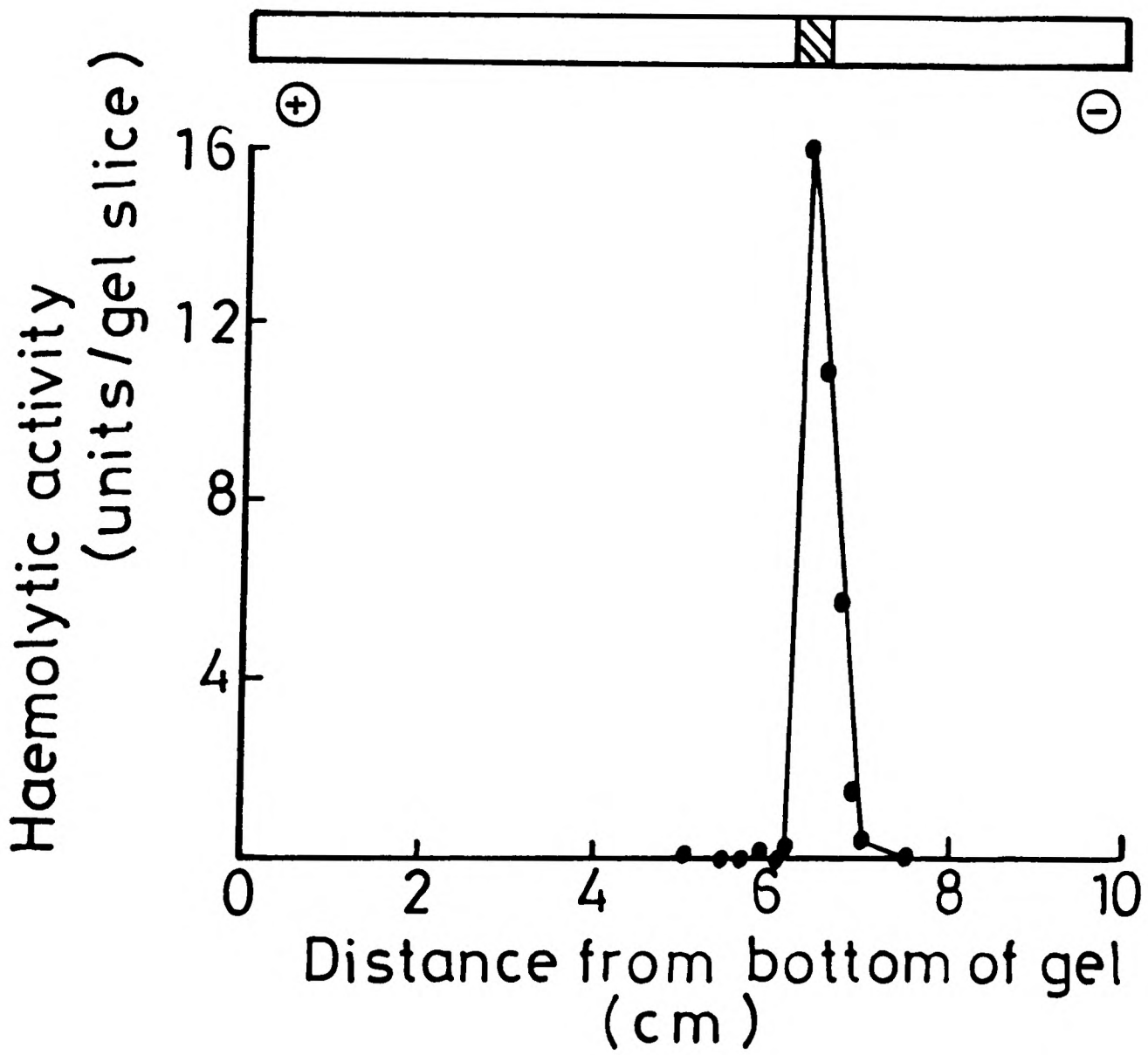
(b) Haemolytic activity

Factor D (10 µg) from the CM-cellulose 32 column was run on an SDS/15% (w/v) polyacrylamide slab gel (section 2.2.2.1). The sample was not reduced or alkylated but was loaded onto the gel in 25% (v/v) glycerol. After electrophoresis, factor $\bar{D}$ was eluted from gel slices (approx. 0.1-0.2 cm) as described in section 2.2.2.1, and the eluates were assayed for haemolytic activity. The Coomassie blue staining of the gel is shown diagramatically above the profile of haemolytic activity.

(a)


Non-reduced
Reduced and alkylated
a
a
b
b
c
c
1
2

(b)


⊕
⊖
Haemolytic activity (units/gel slice)
16
12
8
4
0
2
4
6
8
10
Distance from bottom of gel (cm)

the protein from gel slices of SDS/15% (w/v) polyacrylamide slab gels (Fig. 3.5 (b)).

The purification procedure described gave a final yield of about 0.2 mg of factor $\bar{D}$/1000 ml of serum and a recovery of about 20% of the initial factor $\bar{D}$ haemolytic activity (Table 3.1). The overall purification from serum was approximately 60,000-fold. The yield of factor $\bar{D}$ was consistent with a serum concentration of approximately 1.0-1.5 µg/ml, which is similar to the estimates of Lesavre et al (1979) but slightly lower than suggested by an earlier report (Lesavre and Müller-Eberhard, 1978).

3.3 STABILITY OF FACTOR $\bar{D}$

Factor $\bar{D}$ in 0.1 M Tris/HCl/0.2 M NaCl/2 mM EDTA, pH 8.0, containing 0.02% (w/v) NaN_3, was stable at 4°C for several months with only a slight decrease in haemolytic activity. However, on storage at this temperature, protein aggregation has been observed.

Storage at -70°C prevents aggregation, but freezing and thawing was found to cause a loss of approximately 30% of the original haemolytic activity.

3.4 MOLECULAR WEIGHT OF FACTOR $\bar{D}$

The apparent molecular weight of factor $\bar{D}$ was determined on 15% (w/v) polyacrylamide gels containing 1% (w/v) SDS. In non-reducing conditions the molecular weight was determined as 24,000 and in reducing conditions as 23,700. This estimate is similar to values obtained by others (Volanakis et al, 1977; Lesavre et al, 1979; Davis et al, 1979b). Reduced and alkylated factor $\bar{D}$ had a slightly slower mobility on SDS/polyacrylamide gels, compared with non-reduced factor $\bar{D}$ (Fig. 3.5 (a)). This has also been observed by Volanakis et al (1977).

The estimate of molecular weight obtained under the conditions described is usually within 10% of the values obtained by other techniques (Weber and Osborn, 1975) except that glycoproteins, which show impaired SDS-binding, may behave abnormally and so give erroneous results (Grefrath and Reynolds, 1974). It has not yet been shown conclusively whether factor $\bar{D}$ is a glycoprotein or not (section 3.5).

3.5 CARBOHYDRATE ANALYSIS

3.5.1 Periodic acid/Schiff staining

Factor $\bar{D}$ was stained for reducing sugars using the periodic acid/Schiff (PAS) reaction (Kapitany and Zebrowsky, 1973). Three separate samples containing 20-30 µg of factor $\bar{D}$ were stained with PAS stain following SDS/15% (w/v) polyacrylamide gel electrophoresis. The results obtained for factor $\bar{D}$, the glycoprotein retained by Concanavalin A-Sepharose and control proteins are summarised in Table 3.2. At least 0.3 µg of carbohydrate could be readily detected by this method. Since 20-30 µg of factor $\bar{D}$ was used, this suggested that it contained <1.5% (w/w) reducing sugars. This is in contrast to results obtained by both Davis *et al* (1979b) and Lesavre *et al* (1979) and suggests that factor $\bar{D}$ prepared by their methods may also contain a contaminating glycoprotein.

3.5.2 Hexosamine analysis

Hexosamine was estimated using p-toluene sulphonic acid (section 2.2.9.1). Control samples of lysozyme containing added glucosamine and galactosamine were cohydrolysed but repeatedly gave disappointingly low yields (50% for glucosamine; 30% for galactosamine). The reason for these low yields was not apparent, as under identical conditions, Allen and Neuberger (1975) obtained yields of about 80%. Duplicate samples of factor $\bar{D}$, after correction for low yields, contained <0.5 nmoles/nmole factor $\bar{D}$ for both glucosamine and galactosamine. Although this suggested that factor $\bar{D}$ does not contain significant amounts of hexosamine, the result requires confirmation using a larger sample size.

Analyses for neutral sugars and sialic acid were not carried out.

Table 3.2 Periodic acid/Schiff stain for reducing sugars

Protein	% (w/w) carbohydrate†	µg used	staining	µg carbohydrate present
Controls:				
C1q (A+B chains)	average 10%	2.9	+	0.29
C1q (C chain)	4%	2.9	+	0.12
C1q (C chain)	4%	1.4	-	0.06
$C\bar{1}$ INH	35%	1.0	+	0.35
IgG	3%	10	+	0.30
Ovalbumin	3%	10	+	0.30
Chymotrypsinogen	0%	10	-	0
Bovine serum albumin	0%	10	-	0
25-28 mS protein	?	20	+	N.D.
Factor $\bar{D}$	?	25	-	N.D.

Samples (except for IgG) were reduced and alkylated before electrophoresis on SDS/15% (w/v) polyacrylamide slab gels (Laemmli, 1970). IgG was alkylated without reduction. The gel was stained with Schiff stain after periodic acid oxidation (Kapitany and Zebrowsky, 1973).

N.D. = not determined

†IgG, ovalbumin: Sober, 1970; C1q: K.B.M. Reid (unpublished results); $C\bar{1}$ INH: Haupt *et al*, 1970.

3.6 AMINO ACID COMPOSITION

The amino acid composition calculated from 10 different preparations (Table 3.3) showed no unusual features and was in good agreement with that reported by Davis *et al*, 1979b (see Table 1.4). However, it differed significantly from an earlier report by Volanakis *et al* (1977) in which ile was reported as 6.0 residues/100 residues compared with 2.5 residues/ 100 residues in this study; alanine was 5.8 compared with 8.7; cystine was 2.9 compared with 3.9; leucine was 6.9 compared with 10.2; and lysine was 7.1 compared with 3.8.

Tryptophan was analysed using the method of Penke *et al* (1974). Samples of lysozyme cohydrolysed with factor $\bar{D}$ consistently gave a 75% yield of the theoretical value. This was lower than the yield obtained by Penke *et al* under identical conditions (~95%). Assuming a 75% recovery of factor $\bar{D}$-tryptophan, the estimated tryptophan content was 2.7 nmoles/nmole of factor $\bar{D}$.

Table 3.3 Amino acid composition of factor $\bar{D}$

Amino acid	Composition expressed as residues/100 residues	mol/mol
Asp	9.3	19.9
Thr	3.5	7.5
Ser	4.9	10.5
Glu	9.3	19.9
Pro	8.0	17.1
Gly	10.3	22.0
Ala	8.7	18.6
Val	7.9	16.9
Cys	3.9	8.3
Met	1.2	2.6
Ile	2.5	5.4
Leu	10.2	21.8
Tyr	1.8	3.9
Phe	1.3	2.8
His	5.1	10.9
Lys	3.8	8.1
Arg	7.0	15.0
Trp	1.3	2.7

The composition was calculated from 10 preparations of factor $\bar{D}$. Samples (approx. 1 nmole) were hydrolysed for 48 h at 110^{o}C. No corrections were made for serine and threonine destruction. Cystine was measured as SCM-cysteine and tryptophan was measured using mercaptoethanesulphonic acid. The composition expressed as mol or residue/mol of factor $\bar{D}$ was calculated assuming a mol. wt. of 24,000 for factor $\bar{D}$ and 112 for an amino acid.

3.7 INVESTIGATIONS FOR A POSSIBLE RELATIONSHIP BETWEEN FACTOR $\bar{D}$ AND C1 SUBCOMPONENTS

3.7.1 Immunochemical cross-reactivity with $C\bar{1}s$

Crude factor $\bar{D}$ (approximately 12 ng in 6 µl) obtained by fractionating serum on Sephadex G-75 was incubated for 20 h at 4°C with $F(ab')_2$ anti-$C\bar{1}s$ antiserum (0-8 µl), and the factor $\bar{D}$ haemolytic activity measured. The $F(ab')_2$ antisera had no effect on haemolytic activity.

3.7.2 Investigations for collagen-like regions in factor $\bar{D}$

Factor $\bar{D}$ has a similar apparent molecular weight to the individual chains of the C1q molecule (Reid et al, 1972). Since in the early stages of developing a purification method for factor $\bar{D}$, a high glycine level was observed in the amino acid composition of partially purified factor $\bar{D}$ (>15 residues/100 residues) the possibility that factor $\bar{D}$ may have some collagen-like structure was investigated.

(1) Immunochemical cross-reactivity with C1q

By double diffusion in Ouchterlony plates containing agarose, it was shown that a protein which is possibly a fragment of C1q copurifies with Factor $\bar{D}$ to the Sephadex G-75 stage of purification. The fragment did not bind to Concanavalin A-Sepharose and it was lost in the initial wash on CM-cellulose 32. No immunochemical cross-reactivity of purified factor $\bar{D}$ with C1q was observed.

(2) Collagenase digestion of factor $\bar{D}$

Crude factor $\bar{D}$ (75 ng) was incubated with collagenase (section 2.2.8.2) and assayed for retention of haemolytic activity. No decrease in the haemolytic activity was observed following incubation with as much as 10 ng of collagenase.

(3) Hydroxyproline and hydroxylysine analysis

No hydroxyproline or hydroxylysine was found in factor $\bar{D}$.

3.7.3 Conclusions

Factor $\bar{D}$ appears to play a similar role in the alternative pathway to that of $C\bar{1}s$ in the classical pathway. In cleaving C3-bound factor B, factor $\bar{D}$ causes activation of the alternative pathway C3 convertase in an analagous manner to which $C\bar{1}s$, in cleaving C2, activates the classical pathway C3 convertase, $C\overline{42}$.

It was shown by Müller-Eberhard and Götze (1972) that factor $\bar{D}$ shows no immunochemical or functional cross-reactivity with $C\bar{1}s$. The absence of inhibition of factor $\bar{D}$ haemolytic activity by anti-C1s antisera was consistent with this earlier observation.

As shown in Table 3.3, the earlier finding of >15 glycine residues/100 residues of factor $\bar{D}$ was erroneous and could have been caused by one of the contaminants which was later removed by CM-cellulose 32 and/or Concanavalin A-Sepharose. Indeed, the finding in the initial wash on CM-cellulose 32 of a protein which was immunochemically related to C1q supported this conclusion. In view of these findings, the lack of hydroxyproline and hydroxylysine and the resistance of factor $\bar{D}$ to collagenase digestion were not unexpected.

3.8 EFFECT OF Ca^{2+} ON GEL FILTRATION PROPERTIES

The electrophoretic mobility of factor $\bar{D}$ in agarose has been found to vary depending on whether Ca^{2+} or Mg^{2+} ions are present or not (Davis *et al*, 1979b). To see if Ca^{2+} ions affected the gel filtration properties of factor $\bar{D}$, one sample of purified factor $\bar{D}$ (1 mg) was chromatographed on a column (1.5 cm x 100 cm) of Sephadex G-75 super-fine equilibrated with 0.1 M Tris/HCl/0.2 M NaCl, pH 8.0 (at $4^{o}C$) containing 5 mM EDTA, and a second sample was chromatographed on a similar column equilibrated with the same buffer but containing 2.5 mM $CaCl_2$ instead of EDTA. Molecular weight markers (bovine serum albumin, chymotrypsinogen and cytochrome c) were also run in both buffers.

The gel filtration properties of factor $\bar{D}$ were independent of the presence or absence of Ca^{2+} ions. Fig. 3.6 shows the elution profile in the presence of Ca^{2+} ions only. However, the elution pattern was identical in the presence of EDTA, and in both cases, only one peak of protein and haemolytic activity was obtained.

The binding of Ca^{2+} ions to bovine trypsin causes a conformational change to a more compact structure (Lazdunski and Delaage, 1965) and also results in increased stability of the enzyme (Radhakrishnan *et al*, 1967). It would be of interest to discover whether such conformational changes with resultant changes in stability occur in factor $\bar{D}$ in the presence of Ca^{2+} ions.

Whether there is a specific binding site on factor $\bar{D}$ for Ca^{2+} or Mg^{2+} ions has yet to be demonstrated.

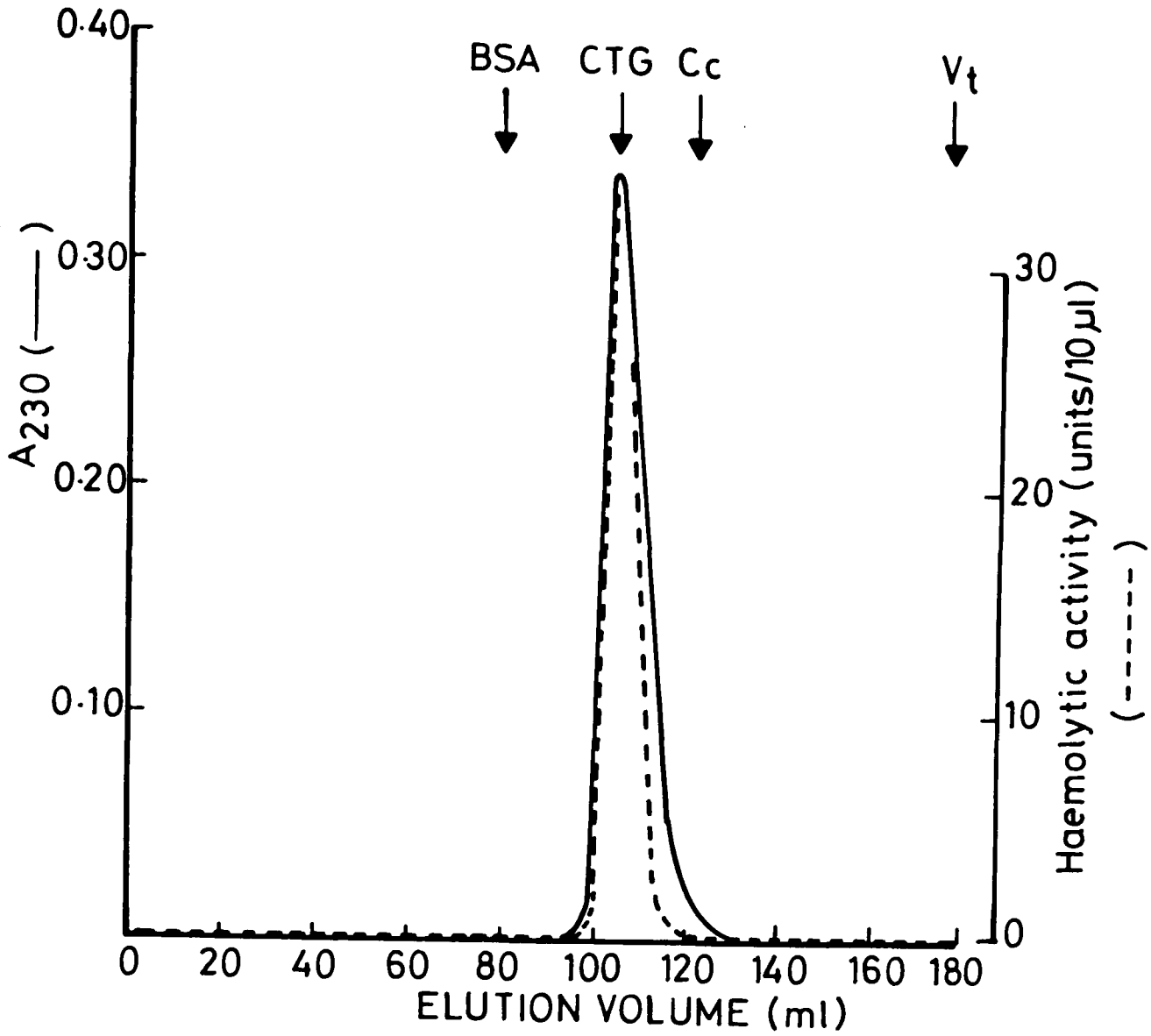


Fig. 3.6 Gel filtration of purified factor $\bar{D}$ on Sephadex G-75 superfine in the presence of Ca^{2+} ions

Purified factor $\bar{D}$ (1 mg) in 0.1 M Tris/HCl/0.2 M NaCl, pH 8.0 (at $4^{o}C$) containing 2.5 mM $CaCl_2$ was applied to a column of Sephadex G-75 superfine equilibrated with the same buffer. Details are given in the text. A similar column equilibrated with the same buffer but containing 5 mM EDTA instead of $CaCl_2$ showed the same elution pattern. Molecular weight markers (with abbreviation and mol.wt. in brackets) were albumin (BSA, 67,000), chymotrypsinogen (CTG, 25,800) and cytochrome c (C_c, 12,400). V_t = total column volume.

CHAPTER 4

AMINO ACID SEQUENCE STUDIES ON FACTOR $\bar{D}$

4.1 INTRODUCTION

At the time of embarking on the sequence studies reported here, no sequence data had been obtained on factor $\bar{D}$ except for the N-terminal amino acid (isoleucine) reported by Volanakis *et al* (1977). Subsequently, two separate reports on the N-terminal sequence have been published. The first reported a tentative sequence of the first nine residues (Davis *et al*, 1979a,b) while the second (Volanakis *et al*, 1980) was a more extensive but still incomplete sequence of the first 57 residues. †

Factor $\bar{D}$ for sequence studies was prepared as described in section 2.2.4.3 and was homogeneous on SDS / 15% (w/v) polyacrylamide gels after reduction and alkylation. Except for the sequencing of the N-terminal of the intact molecule, factor $\bar{D}$ was reduced and alkylated before further modification and/or fragmentation.

The initial strategy used in sequencing factor $\bar{D}$ was to obtain an N-terminal sequence on the intact, unmodified molecule. Subsequent to this, two types of chemical cleavage which produced a small number of peptides were carried out, using CNBr to cleave methionyl residues and in a separate study, o-iodosobenzoic acid to cleave tryptophanyl residues. Carboxypeptidase-Y digestion was also carried out in order to obtain information on the C-terminal residues.

† At the time of writing, Davis (1980) has also published data on the CNBr peptides of factor $\bar{D}$.

Having obtained sufficient basic information, tryptic digestion of succinylated factor $\bar{D}$ yielded a complete set of smaller peptides which could be aligned by homology either with earlier sequence data or with sequence data obtained by others for other serine proteases. Further studies to obtain overlapping sequences which would confirm the sequence deduced by homology included digestion of one of the CNBr peptides with V8 protease. A summary of sequencer results is found in Fig. 4.33.

4.2 N-TERMINAL AMINO ACID SEQUENCE OF INTACT FACTOR $\bar{D}$

The initial sequencer studies carried out on factor $\bar{D}$ were inconclusive because of interference from low molecular weight contaminants and are not reported here. However, the partial sequence of the first 36 residues was established using a 60 nmole preparation of factor $\bar{D}$ (Table 4.1, Fig. 4.1). The protein which was not reduced or alkylated was subjected to 40 cycles of Edman degradation.

As expected from earlier work (Volanakis *et al*, 1977) the single N-terminal amino acid of factor $\bar{D}$ was isoleucine. Since the yield of phenylthiohydantoin (PTH) amino acid at the first cycle was low (6.5 nmoles) this made interpretation of data beyond cycle 36 difficult, and therefore data obtained beyond this point is not reported here.

Three residues were not obtained in this sequence. These were residues 22, 26 and 35. The identity of residues 26 and 35 were obtained from later studies (section 4.3 and 4.4) and were identified as half-cystine and tryptophan respectively. The identification of residue 26 as half-cystine was obtained from sequence analysis of CNBr peptides of reduced and alkylated factor $\bar{D}$. The lack of identification of this residue in the intact molecule would therefore be expected since the disulphide bonds were not reduced and alkylated. The tryptophan at residue 35 occurs close to the point at which it was considered that unequivocal identification of the PTH-derivatives could not be made, and thus positive identification was not possible from this run. Residue 22 was not

Table 4.1 Automated sequence analysis of factor D̄

Residue Number	Amino acid identified	Yield (nmoles)
1	Ile	6.5
2	Leu	7.9
3	Gly	10.3
4	Gly	9.2
5	Arg	1.4
6	Glu	7.9
7	Ala	>8.0
8	Glu	5.3
9	Ala	>8.0
10	His	0.8
11	Ala	>8.0
12	Arg	1.3
13	Pro	1.6
14	Tyr	2.0
15	Met	5.8
16	Ala	4.3
17	Ser	0.5
18	Val	6.1
19	Gln	2.0
20	Leu	3.5
21	Asn	0.7
22	-	-
23	Ala	1.9
24	Glu	0.8
25	Leu	2.2
26	-	-
27	Gly	2.1
28	Gly	2.2
29	Val	3.2
30	Leu	2.3
31	Val	1.6
32	Ala	1.4
33	Glu	0.7
34	Gln	0.6
35	-	-
36	Val	1.7

Automated Edman degradation was carried out as described in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. -, not identified.

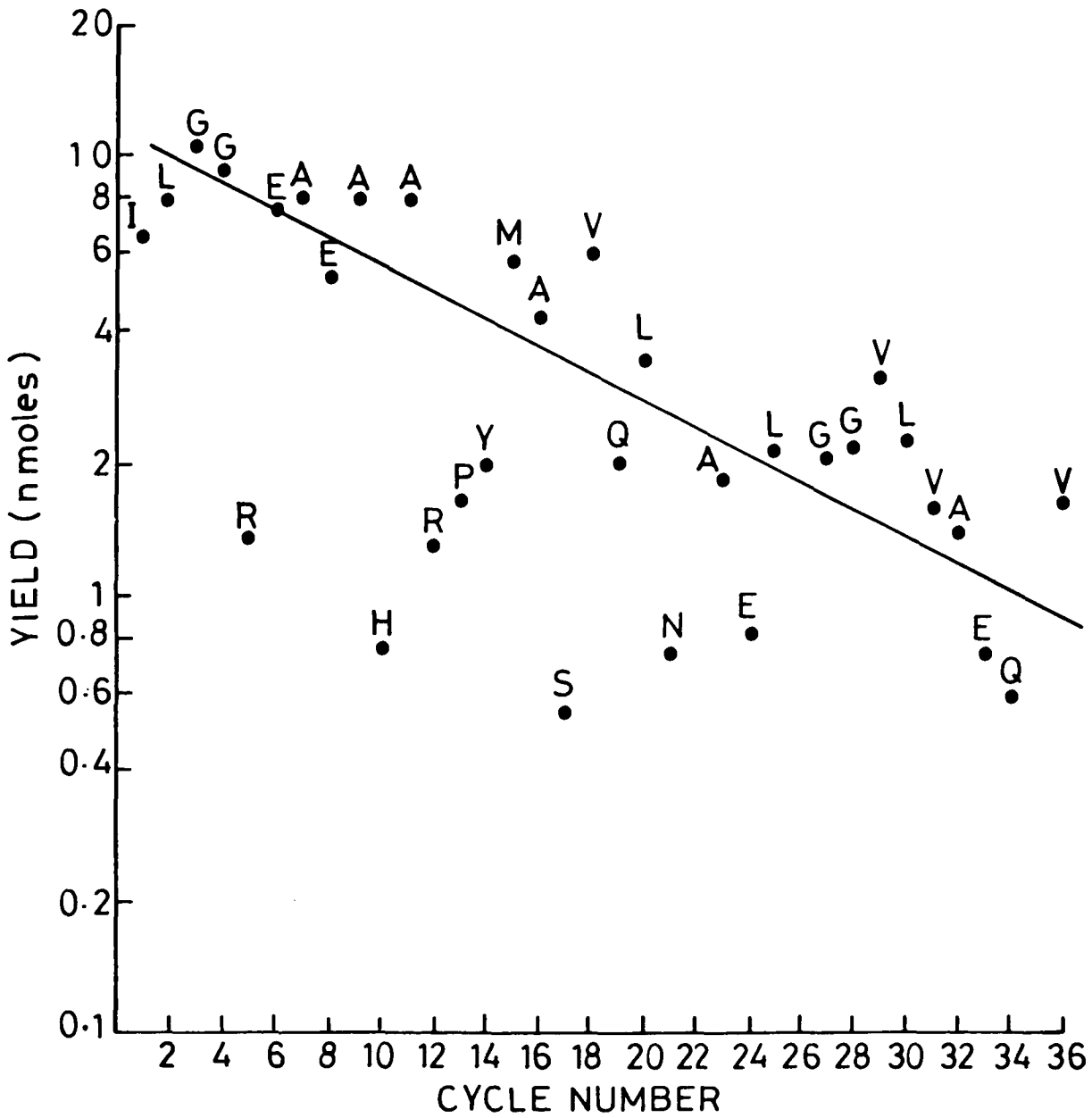


Fig. 4.1 Yield per cycle from automated sequence analysis of intact factor $\bar{D}$

The sequence is denoted by the one letter code for amino acids, and is that of Table 4.1. The amount used in the sequencer was 60 nmoles. The stepwise yield calculated from leu 2, 20, 25 and 30 was 93%.

unequivocally identified. Volanakis et al (1980) found glycine in this position. Although glycine was present at this point in the runs there was no significant increase above background, even although glycine was later identified unequivocally at residues 27 and 28. Glycine had been present at this point in earlier sequencer runs, but could not be unambiguously identified because of the presence of contaminating sequences, and the high background of other PTH-amino acids.

The N-terminal amino acid sequence reported here is in exact agreement with the identification of residues 1-5 reported by Davis et al (1979a,b) and also is in agreement with the identification of positions 6-9, also reported by Davis et al who were unable to distinguish between glutamic acid and glutamine, and between alanine and serine in the procedure they used for identifying the amino acid PTH derivatives.

However, the amino acid sequence differs at several points from that obtained by Volanakis et al (1980) in which phenylalanine was reported at residue 10 compared with histidine in this study; valine at residue 15 compared with methionine; and histidine at residue 24 compared with glutamic acid.† In all three cases the amino acids reported in this study were obtained in yields which were consistent with their being the major amino acid at their respective cycles (Fig. 4.1). The presence of methionine at residue 15 was also confirmed by analysis of the CNBr digest of factor $\bar{D}$ (section 4.3.3) which contained a peptide

† Davis (1980) has also found histidine at residue 24.

with the same amino acid composition as the N-terminal 15 residues reported here. The amino acid sequence reported by Volanakis *et al* also contained two unidentified residues at positions 17 and 26. However, the method used by Volanakis *et al* to detect PTH amino acids did not identify the PTH derivatives of serine or half-cystine.

4.3 CLEAVAGE AT METHIONYL BONDS USING CYANOGEN BROMIDE

4.3.1 Introduction

Cyanogen bromide will selectively cleave peptide bonds on the carboxyl side of methionine residues (Gross and Witkop, 1962). The reaction involves cleavage of the thioether bond of methionine under strongly acidic conditions, and the formation of a transient homoserine iminolactone. This is rapidly hydrolysed to release a peptide with a new N-terminal amino acid and a modified peptide having homoserine or homoserine lactone as C-terminal. Cyanogen bromide will also react with cysteinyl sulphur but the reaction can be prevented by prior carboxymethylation of the thiol group.

Since methionine is a relatively rare amino acid in proteins, digestion with cyanogen bromide usually produces a small number of relatively large peptides which after purification can be sequenced directly or subjected to further degradative treatments (section 4.6). Factor $\bar{D}$ contains 1.2 methionines/100 amino acid residues. Assuming a molecular weight of approximately 24,000, three or possibly four peptides were expected following cleavage with cyanogen bromide.

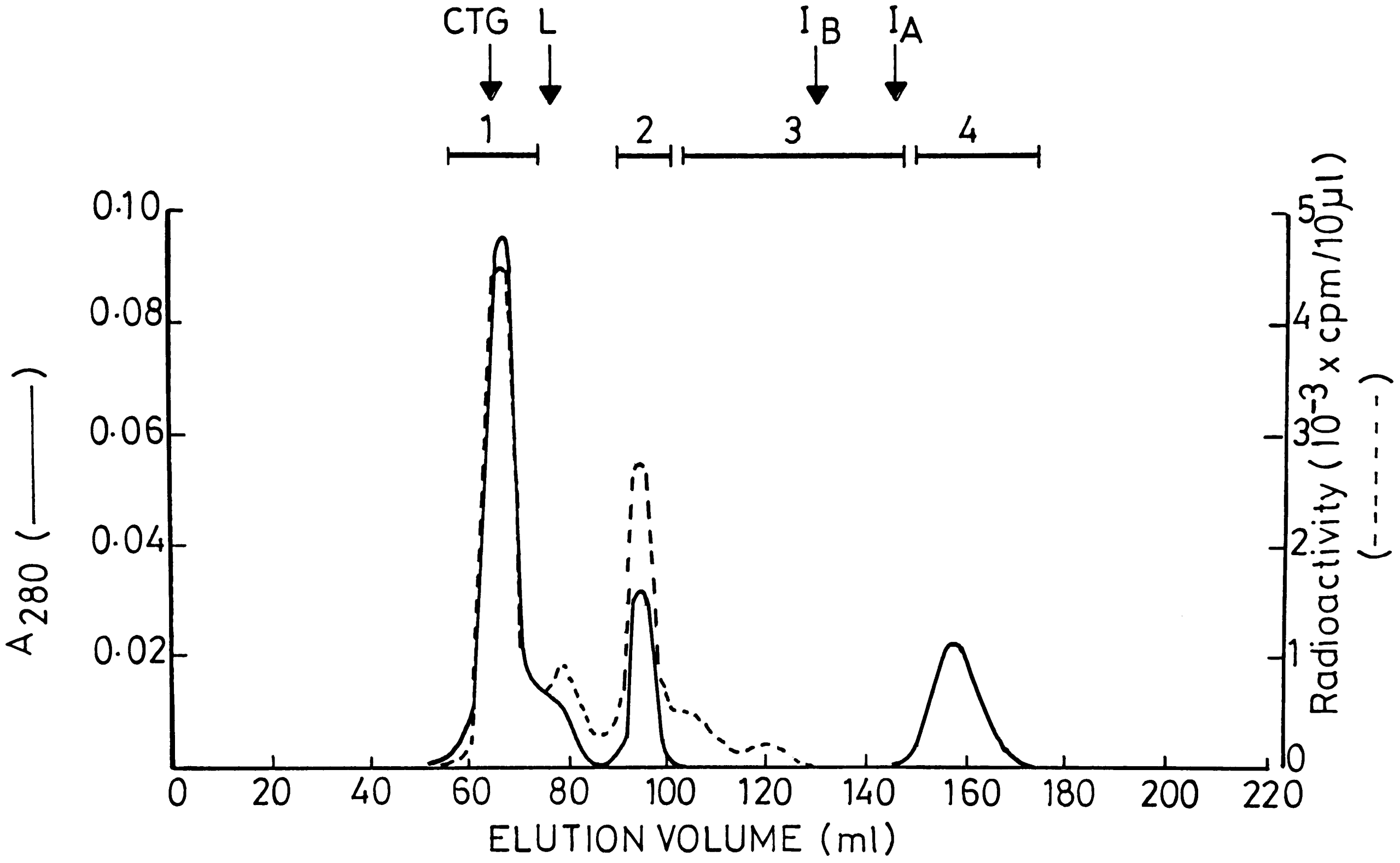
4.3.2 CNBr digestion and purification of peptides

4.3.2.1 50 nmole digest labelled with iodo-[2-^{14}C]-acetic acid

Factor $\bar{D}$ was reduced, alkylated and labelled with iodo-[2-^{14}C]-acetic acid and then digested with CNBr (section 2.2.11.1). The digest was initially fractionated on Sephadex G-50 superfine (Fig. 4.2). Two major peaks of protein containing radioactivity eluted at 56-76 ml and at 90-104 ml respectively.

Fig. 4.2 Purification on Sephadex G-50 superfine of peptides obtained by CNBr digestion of factor D labelled with iodo-[2-^{14}C]-acetic acid

Peptides were eluted using 5% (v/v) formic acid. Details are given in section 2.2.11.1. The following markers were reduced and alkylated and used to calibrate the column (abbreviation and mol. wt. in brackets): chymotrypsinogen (CTG, 25,700), lysozyme (L, 14,300), insulin B chain (I_B, 3,400) and insulin A chain (I_A, 2,400). The column fractions were pooled as indicated by bar-lines. Peptides CNBr-1, CNBr-2 and CNBr-3 were isolated from pools 1, 2 and 3 respectively.


CTG
L
I B
I A
1
2
3
4
0·10
0·08
0·06
0·04
0·02
A280 (——)
0
20
40
60
80
100
120
140
160
180
200
220
ELUTION VOLUME (ml)
5
4
3
2
1
0
Radioactivity (10-3 x cpm/10µl)
(- - - - -)

The second peak (90-104 ml) contained peptide CNBr-2. The first peak (56-76 ml) was found to be heterogeneous on SDS/20% (w/v) polyacrylamide gels and was repurified on a column (1.5 cm x 100 cm) of Sephadex G-75 superfine (Fig. 4.3 a and b). Peptide CNBr-1 eluted between 53-63 ml on this column. Peptide CNBr-3 was purified from the pool made between 105-150 ml on the initial fractionation of the digest on Sephadex G-50 superfine (Fig. 4.2). This pool was freeze-dried and then redissolved in 10 mM NH_4HCO_3/5% (w/v) CH_3CN (50 µl) plus 99% formic acid (10 µl). The material in the pool was then further purified by HPLC using a µBondapak C-18 column. Two significant peaks of peptide material were detected in an A_{210} ratio of approximately 10:1. The amino acid compositions of these peptides were identical, suggesting that they were two forms of the same peptide. It is possible that one of the peptides had its C-terminal homoserine in the lactone form.

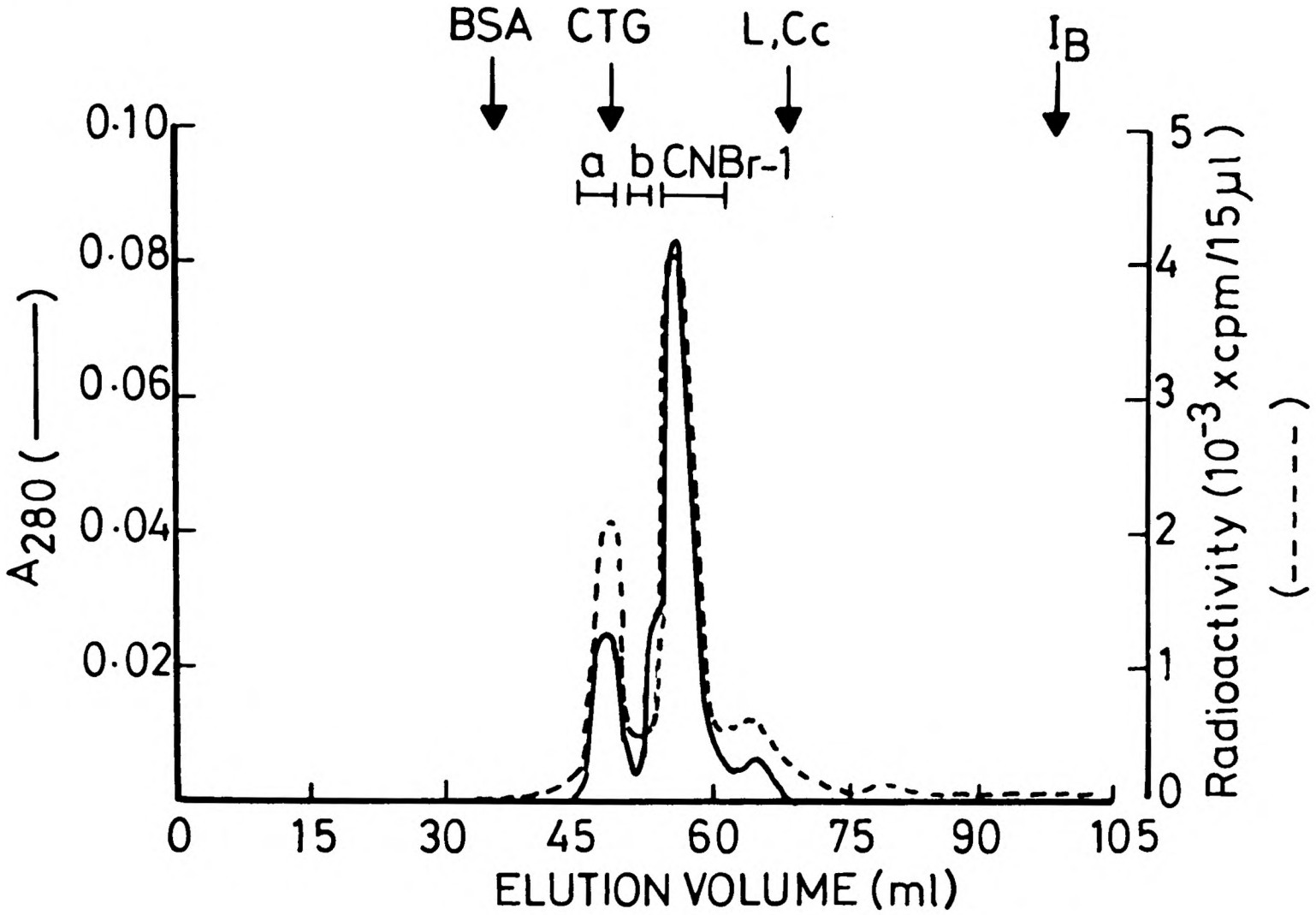
The apparent molecular weights of peptides CNBr-1 and CNBr-2, as estimated on SDS/20% (w/v) polyacrylamide gels, were 15,800 and 6,600 (Fig. 4.4). Peptide CNBr-3 was not detected on SDS/20% (w/v) polyacrylamide gels, but its molecular weight was estimated to be 1700 from its amino acid composition (Table 4.2) and its amino acid sequence (Table 4.3 and Fig. 4.8). The combined molecular weights of these three major CNBr peptides thus account for the total apparent molecular weight of factor $\bar{D}$. Although the possible occurrence of a fourth, very small CNBr peptide has not been formally excluded, when the lower molecular weight fractions from the Sephadex G-50 superfine column were electrophoresed

Fig. 4.3 Purification on Sephadex G-75 superfine of fraction-1 from Sephadex G-50 superfine column

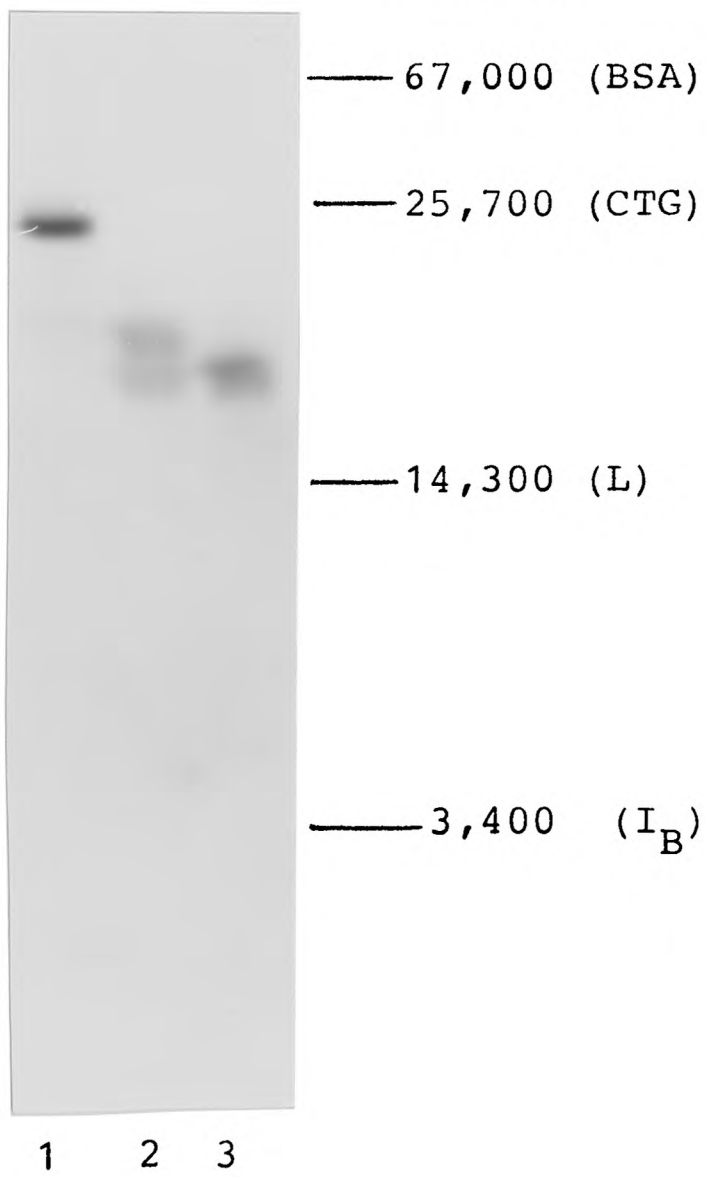
(a) Peptides were eluted in 10% (v/v) formic acid. Details are given in section 2.2.11.1. The following markers were reduced and alkylated and used to calibrate the column (abbreviation and molecular weight in brackets): bovine serum albumin (BSA, 67,000), chymotrypsinogen (CTG, 25,700), lysozyme (14,300), cytochrome c (C_C, 12,400) and insulin B chain (I_B, 3,400). Fractions were pooled as indicated by bar-lines.

(b) SDS/20% (w/v) polyacrylamide gel electrophoresis of 2-5% of the total volume of pools from Sephadex G-75 superfine: the positions and mol. wts. of reduced and alkylated markers are indicated. Tracks 1, 2 and 3 correspond to pools a and b, and CNBr-1 respectively.

(a)


BSA
CTG
L,Cc
I_B
a
b
CNBr-1
0·10
0·08
0·06
0·04
0·02
A_{280} (——)
5
4
3
2
1
0
Radioactivity (10^{-3} xcpm/15µl)
(- - - - -)
0
15
30
45
60
75
90
105
ELUTION VOLUME (ml)

(b)


Mol. wt.
67,000 (BSA)
25,700 (CTG)
14,300 (L)
3,400 (I_B)
1
2
3

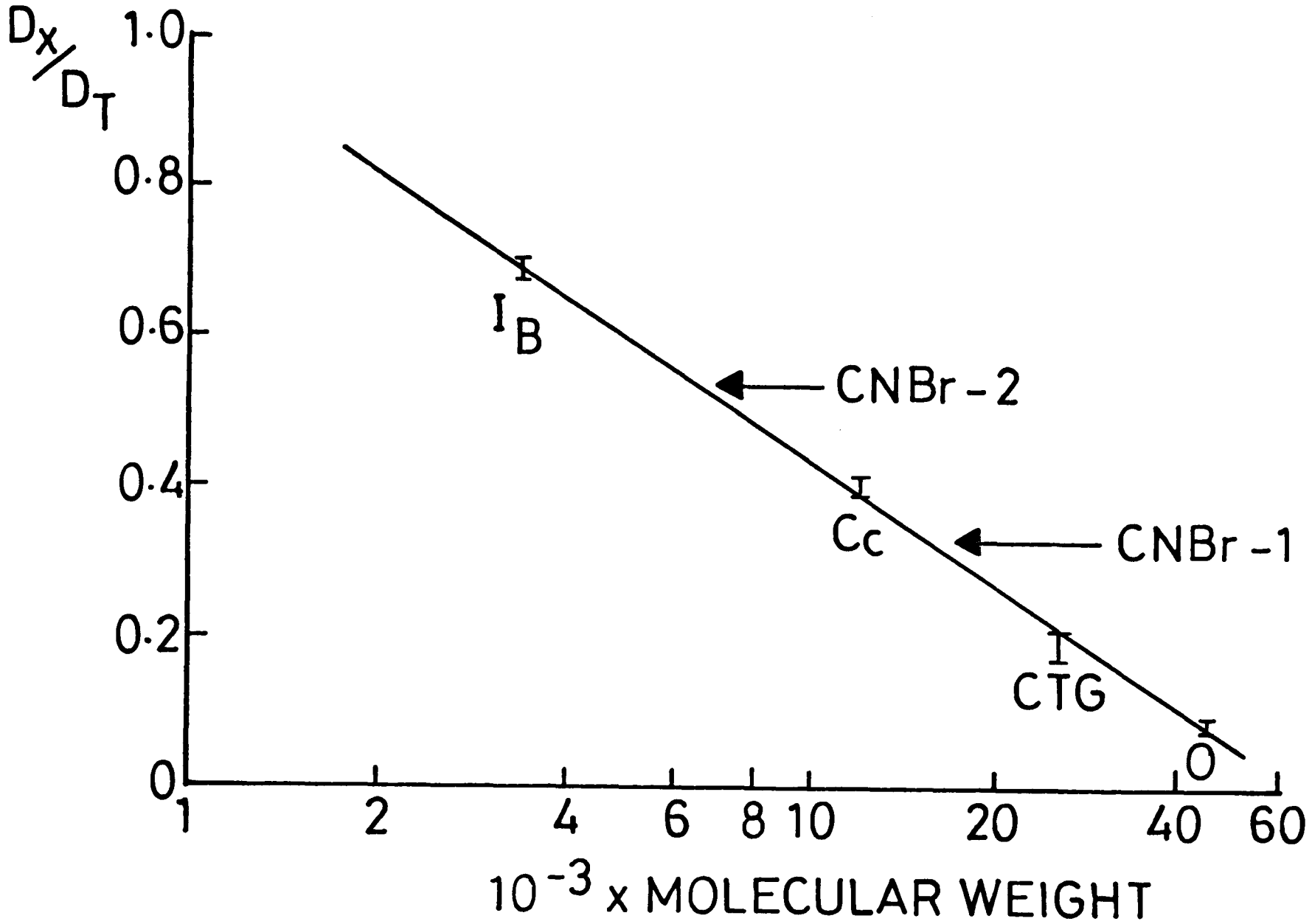


Fig. 4.4 Molecular weight estimation of CNBr peptides by polyacrylamide gel electrophoresis

Electrophoresis of peptides and markers was carried out on SDS/20% (w/v) polyacrylamide gels as described in section 2.2.2.1. D_X/D_T is the relative mobility of a protein or peptide. D_X and D_T were the mobility of protein or peptide, and the mobility of the bromophenol blue marker respectively. Reduced and alkylated protein standards (abbreviation and mol. wt. in brackets) used were ovalbumin (O, 45,000), chymotrypsinogen (CTG, 25,700), cytochrome c (C_c, 12,400) and insulin B chain (3,400).

Table 4.2 Amino acid compositions of CNBr peptides

Amino Acid	Amino acid composition of		
	CNBr-1 (residues/100 residues)	CNBr-2	CNBr-3 (mol / mol)
Asp	10.6	10.0	0.3 (0)
Thr	3.7	3.5	0.3 (0)
Ser	5.4	7.6	-
Hse	0.6	-	0.7 (1)
Glu	9.8	7.9	2.1 (2)
Pro	7.3	6.3	1.1 (1)
Gly	10.2	14.3	2.2 (2)
Ala	10.6	9.0	2.7 (3)
Val	8.3	9.8	0.3 (0)
Cys	1.6	3.0	-
Ile	2.2	4.3	0.8 (1)
Leu	14.1	6.8	1.0 (1)
Tyr	0.4	2.4	0.9 (1)
Phe	0.2	0.7	-
His	6.5	1.9	1.0 (1)
Lys	2.1	4.9	-
Arg	6.4	7.6	1.6 (2)

Details of amino acid analysis are described in sections 2.2.9.1 and 2.2.9.2. Cystine was estimated as S-carboxymethyl cysteine. Tryptophan was not determined, and no corrections were made for destruction of serine and threonine.

at pH 1.9 and the ninhydrin- and starch-iodide-positive spots eluted, no free homoserine was found, nor was any other peptide present except for peptide CNBr-3.

Two other fragments were detected in CNBr digests of factor $\bar{D}$. One of these (of apparent molecular weight 18,000) eluted on Sephadex G-75 superfine just ahead of peptide CNBr-1, from which it was not clearly resolved (Fig. 4.3(b), track 2). This peptide, which was present at about 1/5 the concentration of peptide CNBr-1, probably resulted from incomplete digestion of the methionyl bond at position 15 in factor $\bar{D}$ and would be composed of peptides CNBr-3 and CNBr-1. The other fragment, also present in low amounts, had an apparent molecular weight of 23,000. This fragment was distinguishable from residual uncleaved factor $\bar{D}$ (Fig. 4.3(b), track 1) and is probably composed of peptides CNBr-1 and CNBr-2. Amino acid sequence analysis on these two fragments was not performed.

4.3.2.2 50 nmole digest labelled with $[1,3\text{-}^{14}C]\text{-}iPr_2P\text{-}F$

Factor $\bar{D}$ was labelled with $[1,3\text{-}^{14}C]\text{-}iPr_2P\text{-}F$ as described in section 2.2.10.2. An estimate of the amount of $[1,3\text{-}^{14}C]\text{-}iPr_2P\text{-}F$ incorporated was obtained by scintillation counting. Initially, approximately 8% of the factor $\bar{D}$ was labelled. In a similar experiment Davis (1980) incorporated $[^{3}H]\text{-}iPr_2P\text{-}F$ into 3% of the enzyme. The labelled factor $\bar{D}$ was reduced, alkylated with nonradioactive iodoacetic acid, and then digested with CNBr as described above. The amount of factor $\bar{D}$ which still contained the label had been reduced tenfold to 0.8%. This loss of label was probably

caused by the acidic conditions used in the cleavage reaction, as low pH favours the hydrolysis of the isopropyl groups on iPr_2P-F.

The CNBr digest was fractionated on Sephadex G-50 superfine and Sephadex G-75 superfine as described above. Two peaks of protein containing radioactivity were obtained by fractionation of the digest on Sephadex G-50 superfine (Fig. 4.5). The first peak eluted between 53-74 ml and was refractionated on Sephadex G-75 superfine. All the radioactivity of this first peak eluted between 42-50 ml on the G-75 column and contained partially cleaved and intact factor $\bar{D}$. The second peak of radioactivity from Sephadex G-50 superfine eluted between 91-108 ml and contained the peptide, CNBr-2. The uncleaved and partially cleaved factor $\bar{D}$ was redigested with fresh CNBr but no further production of peptide CNBr-2 occurred.

4.3.3 Automated amino acid sequence analysis

The amino acid composition of the peptides CNBr-1, CNBr-2 and CNBr-3 are given in Table 4.2.

4.3.3.1 Peptide CNBr-1

From the apparent molecular weight (15,800) of peptide CNBr-1, it was estimated that the peptide was approximately 141 amino acid residues long. The N-terminal amino acid of peptide CNBr-1 was alanine, and the sequence (Table 4.3; Fig. 4.6) overlapped with that of intact factor $\bar{D}$ (Table 4.1) from residue 16 onwards. The only exception was at residue 11 of peptide CNBr-1, where SCM-cysteine was found compared with an unidentified residue in factor $\bar{D}$. Three amino acid

Fig. 4.5 Purification on Sephadex G-50 superfine of peptides obtained by CNBr digestion of factor $\bar{D}$ labelled with $[1,3\text{-}^{14}C]\text{-}iPr_2P\text{-}F$

Peptides were eluted using 5% (v/v) formic acid. The purification and markers are the same as is described for Fig. 4.2. The column fractions were pooled as indicated by bar-lines. The radioactivity in peak (1) can be accounted for since this was shown by SDS/20% (w/v) polyacrylamide gel electrophoresis to contain intact and partially cleaved factor $\bar{D}$ as well as the peptide CNBr-1.

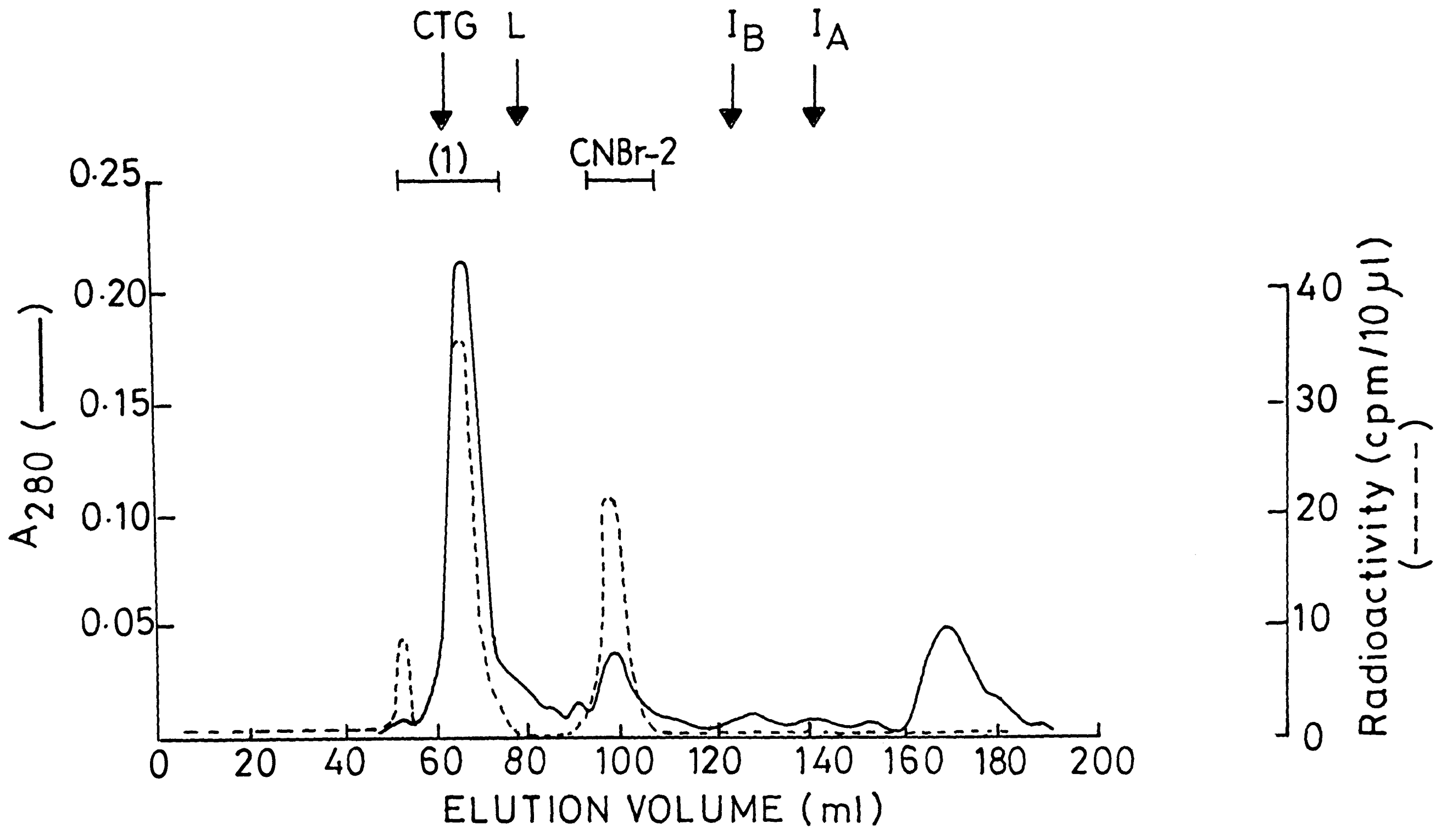

CTG
L
I_B
I_A
(1)
CNBr-2
0·25
0·20
0·15
0·10
0·05
A_{280}
0
20
40
60
80
100
120
140
160
180
200
ELUTION VOLUME (ml)
40
30
20
10
0
Radioactivity (cpm/10 µl)
(----)

Table 4.3 Automated sequence analysis of CNBr peptides

Residue Number	Amino acid identified in Peptide CNBr-1	Peptide CNBr-2	Peptide CNBr-3
1	Ala (3.2)	Cys (10.5)	Ile (6.4)
2	Ser (0.1)	Ala (10.3)	Leu (6.6)
3	Val (3.3)	Glu (7.3)	Gly (5.9)
4	Gln (0.7)	Ser (1.3)	Gly (4.9)
5	Leu (1.6)	Asn (4.1)	Arg (1.2)
6	Asn (0.3)	Arg (0.6)	Glu (1.3)
7	(Gly) (0.3)	Arg (0.7)	Ala (2.7)
8	Ala (0.7)	Asp (3.1)	Glu (0.6)
9	-	Ser (0.8)	Ala (2.0)
10	Leu (0.6)	Cys (5.2)	-
11	Cys	Lys (1.1)	Ala (1.1)
12	Gly (0.2)	Gly (1.4)	
13	Gly (0.4)	Asp (1.5)	
14	Val (0.9)	Ser* (0.6)	
15	Leu (0.7)	Gly (0.7)	
16	Val (1.0)	Gly (0.6)	
17	Ala (0.5)	Pro (0.7)	
18	Glu (0.1)	Leu (2.1)	
19	-	Val (2.5)	
20	-	Cys (2.2)	
21	Val (0.6)	-	
22	Leu (0.5)	-	
23		Val (1.3)	
24		Leu (0.8)	
25		Glu (0.6)	
26		(Gly) (0.5)	

Automated Edman degradation was carried out as described in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields in nmoles are given in brackets. In peptide CNBr-1 the presence of SCM-cysteine PTH derivatives was confirmed by detection of radioactive label. In peptide CNBr-2 labelled with [1-3-^{14}C]-iPr_2P-F, radioactivity was released at residue 14 (indicated by *) as shown in Fig. 4.7. Residues in brackets are only tentatively identified. -, not identified.

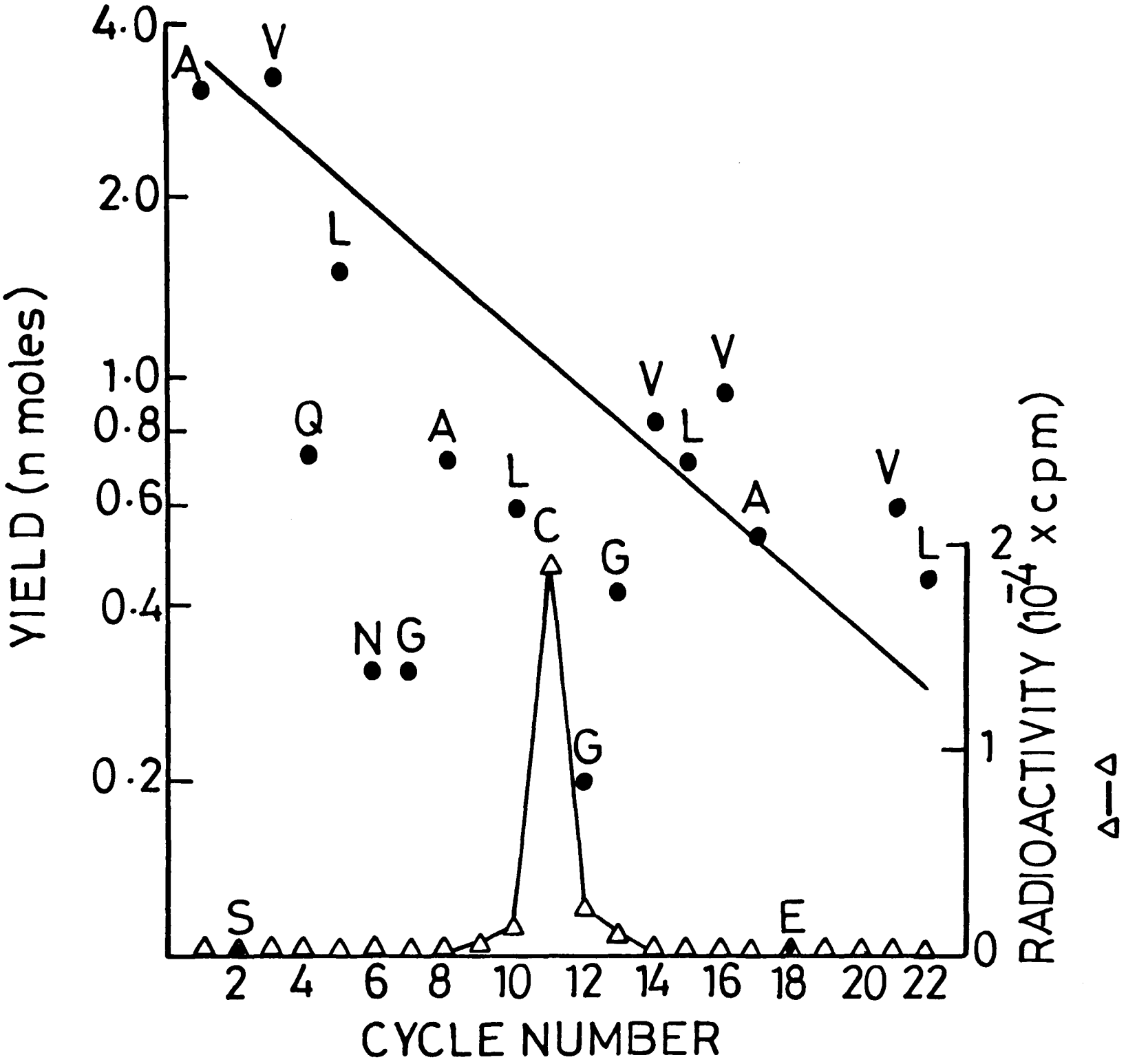


Fig. 4.6 Yield per cycle from automated sequence analysis of peptide CNBr-1

The sequence is denoted by the one letter code for amino acids and is found also in Table 4.3. Radioactivity released (Δ) is from ^{14}C-SCM-cysteine. The amount used in the sequencer was 20 nmoles. The stepwise yield calculated from leu 5, 10 and 15 was 90%.

residues were not identified in the sequencer run on peptide CNBr-1. Of these, two had been detected in the sequencer run of intact factor $\bar{D}$ (corresponding to glutamic acid at residue 9, and glutamine at residue 19 of peptide CNBr-1). The other residue (position 20) was also unidentified in intact factor $\bar{D}$ but was found to be tryptophan in later studies (section 4.4). The identity of residue 7 as glycine in this run was tentative as the yield was low. However, Volanakis et al (1980) and Davis (1980) have found glycine in this position (residue 22 of intact factor $\bar{D}$; cf. section 4.2).

4.3.3.2 Peptide CNBr-2

From both the sequence data obtained for peptides CNBr-1 and CNBr-3, and from the lack of homoserine in the amino acid composition of peptide CNBr-2, it is probable that peptide CNBr-2 is derived from the C-terminal portion of factor $\bar{D}$ (Table 4.3, Fig. 4.7). Peptide CNBr-2 was composed of approximately 60 amino acid residues and automated N-terminal sequence analysis gave the sequence of the first 26 amino acid residues with the exception of residues 21 and 22. From analysis of tryptic peptides of factor $\bar{D}$ (section 4.5) these residues were both identified as glycine. When peptide CNBr-2 was isolated from a CNBr digest of factor $\bar{D}$ that had been labelled with $[1,3\text{-}^{14}C]$-$\text{-}iPr_2P\text{-}F$, it was found to be radiolabelled (Fig. 4.5). By scintillation counting of one-third of the butyl chloride extracts from the sequencer run of peptide CNBr-2 it was shown that the radioactivity in the peptide was located in the serine residue found at position 14 (Fig. 4.7).

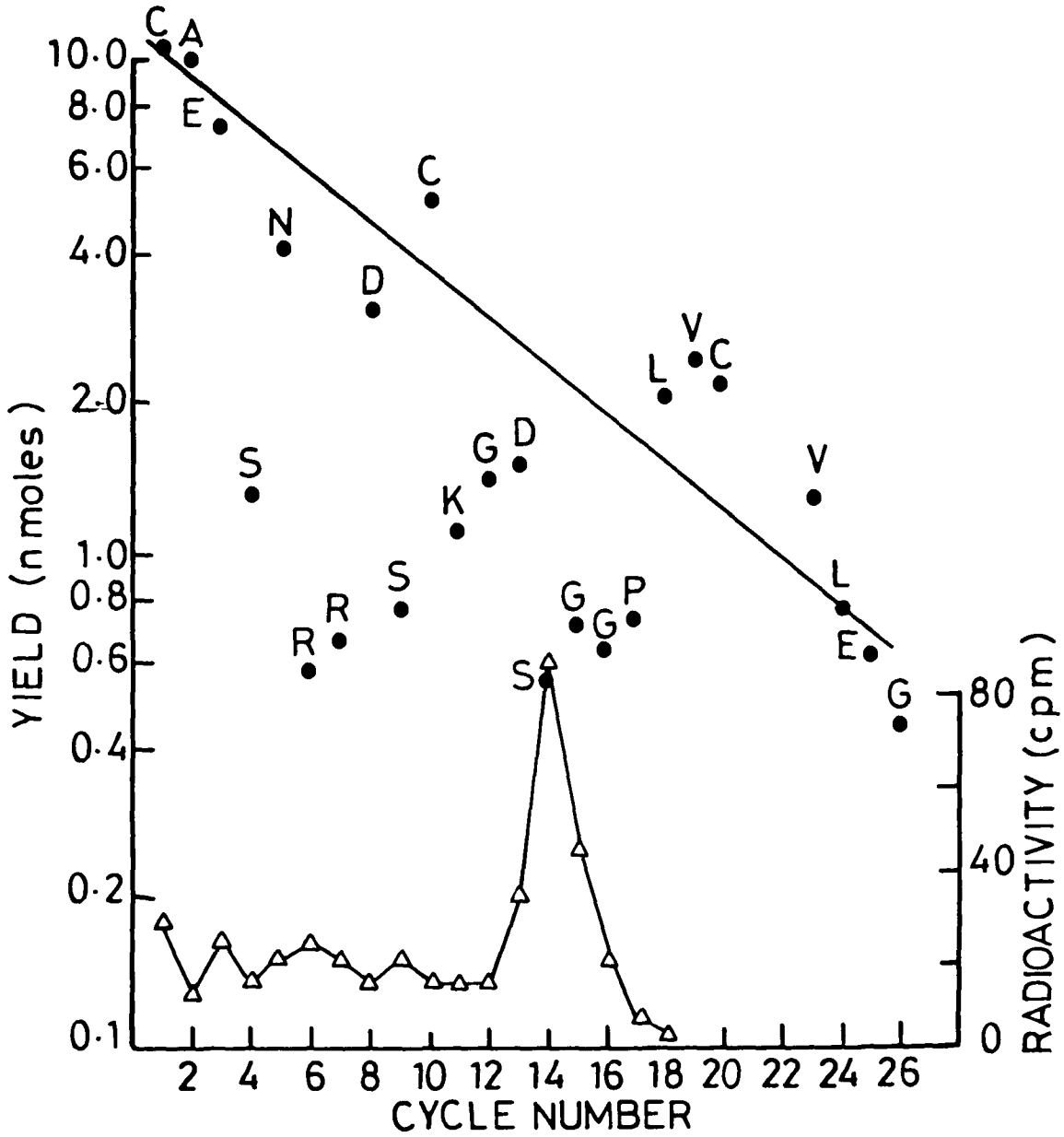


Fig. 4.7 Yield per cycle from automated sequence analysis of peptide CNBr-2

The sequence is denoted by the one letter code for amino acids and is found also in Table 4.3. Radioactivity released (Δ) is from $[1,3\text{-}^{14}C]\text{-}iPr_2P\text{-}F$ label and is associated specifically with residue 14. The amount used in the sequencer was 30 nmoles. The stepwise yield calculated from cys 1, 10 and 20 was 94%.

4.3.3.3 Peptide CNBr-3

The amino acid composition of peptide CNBr-3 (molecular weight 1700) was identical to that of the N-terminal 15 residues of intact factor $\bar{D}$. The N-terminal amino acid sequence was also found to be identical for the first 11 residues except for residue 10 which was not identified in the sequencer run of peptide CNBr-3 (Table 4.3, Fig. 4.8). However, the presence of histidine at residue 10 in intact factor $\bar{D}$ was consistent with the presence of one histidine residue in the amino acid composition of peptide CNBr-3.

4.3.3.4 Discussion

Digestion of factor $\bar{D}$ with CNBr gave three major peptides, CNBr-1, CNBr-2 and CNBr-3 in yields of 26%, 55% and 39% respectively. The yields of peptides CNBr-1 and CNBr-3 were lower than that of CNBr-2 since they each underwent an extra purification step compared with peptide CNBr-2.

About 5% of the original factor $\bar{D}$ remained following cleavage by CNBr. Although the overall cleavage of factor $\bar{D}$ was approximately 95%, cleavage at both methionine residues was not so complete, and small quantities of peptides (probably less than 5 nmoles) comprising either peptides CNBr-3 and CNBr-1 or peptides CNBr-1 and CNBr-2, were present in the digest.

The amino acid sequence data obtained allowed the alignment of the three major peptides as shown in Fig. 4.9. The possibility that there may be a fourth very small peptide has been considered (section 4.3.2) but not formally eliminated,

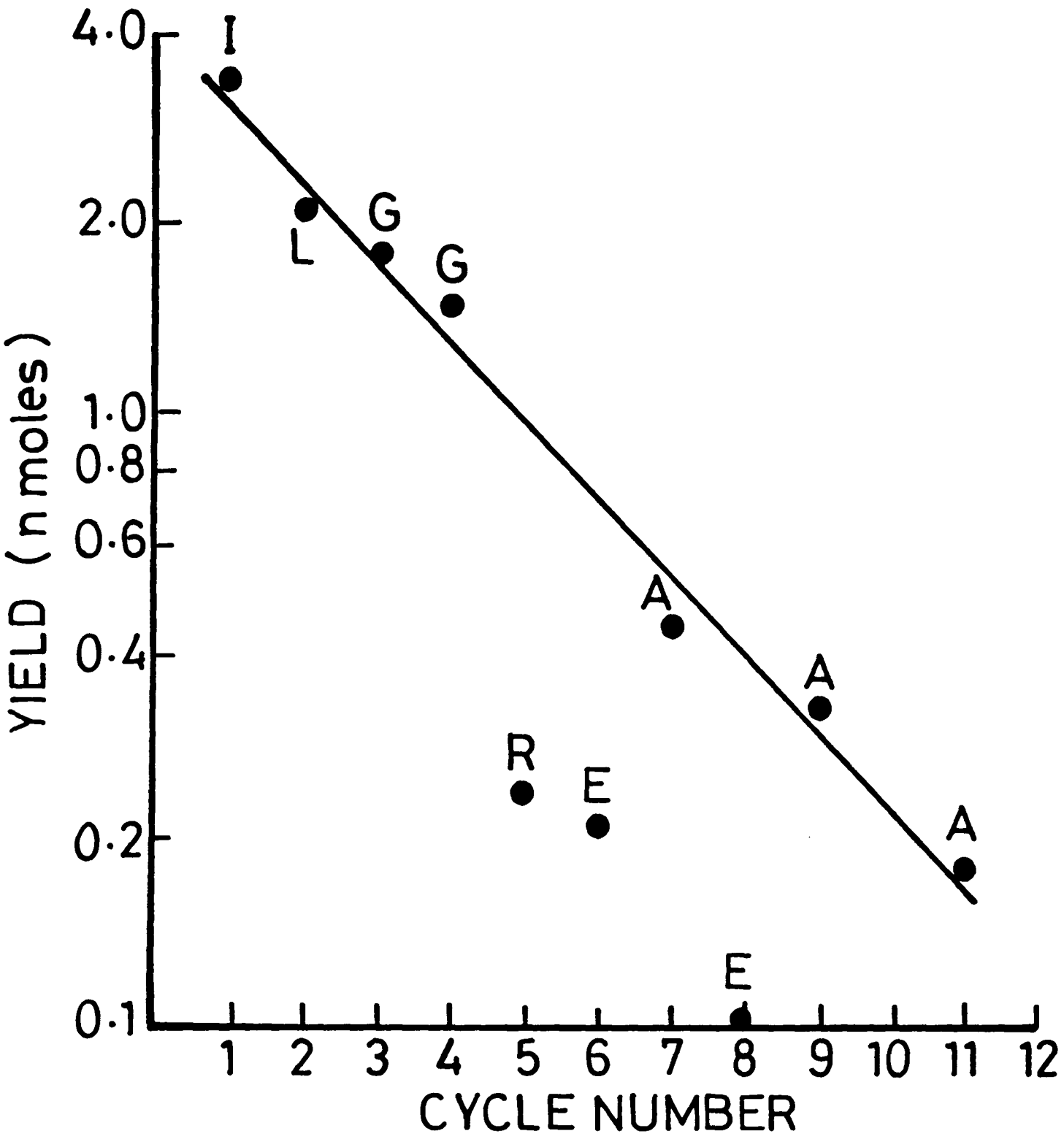


Fig. 4.8 Yield per cycle from automated sequence analysis of peptide CNBr-3

The sequence is denoted by the one letter code for amino acids and is found also in Table 4.3. The amount used in the sequencer was 12 nmoles. The stepwise yield calculated from ala 7,9 and 11 was 75%.

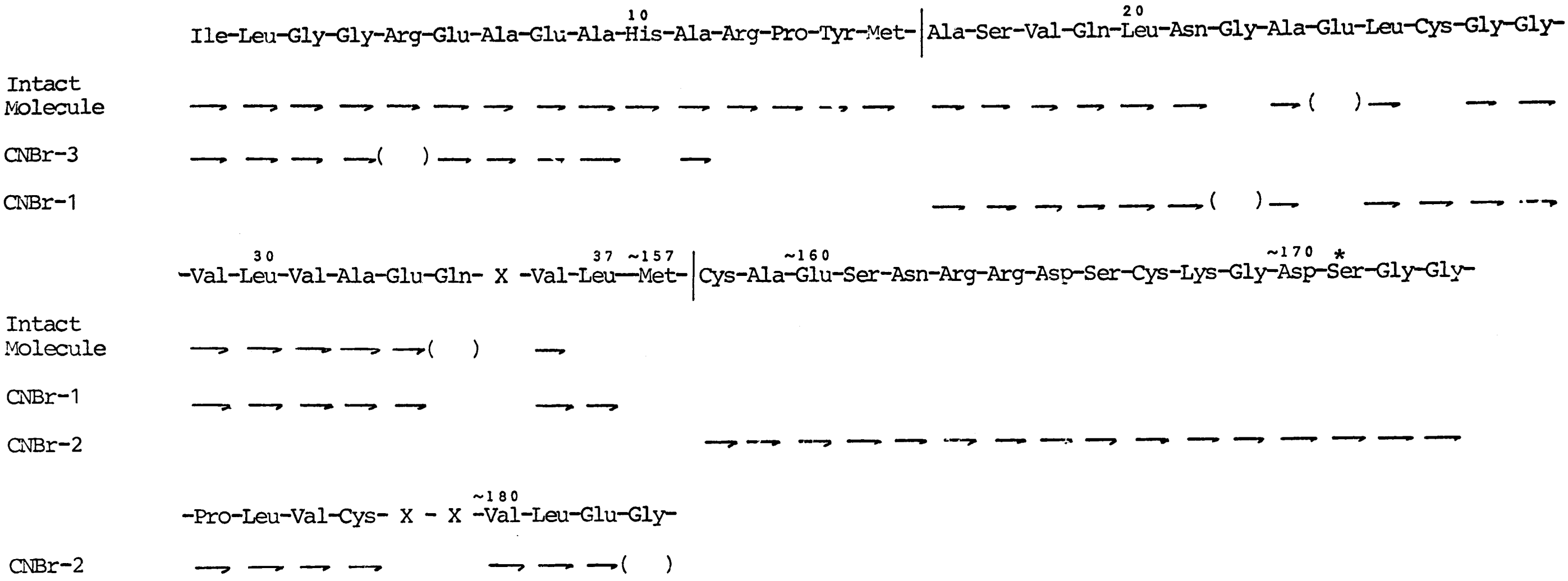


Fig. 4.9 N-Terminal amino-acid sequence of factor D̄ and alignment of the major CNBr fragments

The intact molecule, peptide CNBr-1, peptide CNBr-2 and peptide CNBr-3 have apparent mol. wts. of 24000, 15800, 6600 and 1700 respectively and therefore are composed of approx. 215, 141, 60 and 15 amino acid residues respectively. ⟶, ( ) or a blank space denote positive, tentative or no identification respectively of a particular residue. *Denotes the position of the [1,3-^{14}C]iPr_2P-F labelled serine residue. | denotes supposed cleavage site by CNBr.

since factor $\bar{D}$ contains approximately 2·6 residues of methionine/ molecule (Table 3.3). That any unidentified peptide must be small is suggested from the finding that the apparent molecular weight of factor $\bar{D}$ is virtually accounted for by the sum of the molecular weights of peptides CNBr-1, CNBr-2 and CNBr-3 (15,800, 6,600 and 1,700 respectively).

4.4 CLEAVAGE AT TRYPTOPHANYL BONDS USING O-IODOSOBENZOIC ACID

4.4.1 Introduction

Tryptophan is another amino acid which, like methionine occurs infrequently in proteins. Serine proteases usually contain four or five tryptophan residues which tend to be conserved but are not usually invariant. From the tryptophan analysis of factor $\bar{D}$ (section 3.6) it was estimated that it should contain three tryptophan residues. Thus cleavage at tryptophan using a selective reagent would yield four major fragments.

Fragmentation at trp-X bonds is not generally satisfactory. Cleavage using BNPS-skatole, the most widely used method to date, gives cleavage yields of 15-60% (Omenn et al, 1970). However, the conditions required to obtain more efficient cleavage at trp, also promote modification of methionine, tyrosine and histidine, and cleavage at tyr-X and his-X bonds. Recently, Mahoney and Hermodson (1979) have described a method giving >70% cleavage at tryptophanyl residues using o-iodosobenzoic acid (IBA). Under optimal conditions, methionine was the only other residue to be modified. The mechanism of cleavage, as with BNPS-skatole, is thought to involve oxidative halogenation of the indole ring of tryptophan. This reagent was used to cleave factor $\bar{D}$ in the sequence studies reported here.

4.4.2 Digestion with o-iodosobenzoic acid and purification of peptides

4.4.2.1 Pilot study using 8 nmoles of factor $\bar{D}$

The uncleaved and partially cleaved factor $\bar{D}$ which remained following digestion with CNBr (section 4.3) was

used for the pilot study. The protein was digested with IBA as described in section 2.2.11.2. Electrophoresis on SDS/20% (w/v) polyacrylamide slab gels was not successful but on SDS/12% (w/v) polyacrylamide cylindrical gels at least two bands of apparent molecular weight 11,500 and 10,500 were observed by both Coomassie Blue staining and by scintillation counting of gel slices (Fig. 4.10). Both bands stained with similar intensity with Coomassie Blue. A trace of uncleaved factor $\bar{D}$ was also present. This suggested that factor $\bar{D}$ had been cleaved roughly in half, probably with one or more low molecular weight fragments which were lost either through dialysis or during electrophoresis.

4.4.2.2 70 nmole digest

The IBA digest was initially fractionated on Sephadex G-50 superfine (Fig. 4.11). Two major peaks of protein containing radioactivity eluted at 58-70 ml and 70-91 ml respectively. The first peak, IBA-1 (58-70 ml) was heterogeneous on SDS/20% (w/v) polyacrylamide gels and contained intact factor $\bar{D}$ and a peptide of apparent molecular weight 21,500. The second peak, IBA-2 (70-91 ml) was also heterogeneous on SDS/20% (w/v) polyacrylamide slab gels. This protein peak covered an apparent molecular weight range (estimated from the G-50 column) of between 10,000 and 19,000 (Fig. 4.12). The fraction was repurified on a column of Sephadex G-75 superfine, but the peptides were not resolved further. Samples were taken for SDS/polyacrylamide gel electrophoresis from the ascending and descending sides of the protein peak. Scintillation counting of gel slices

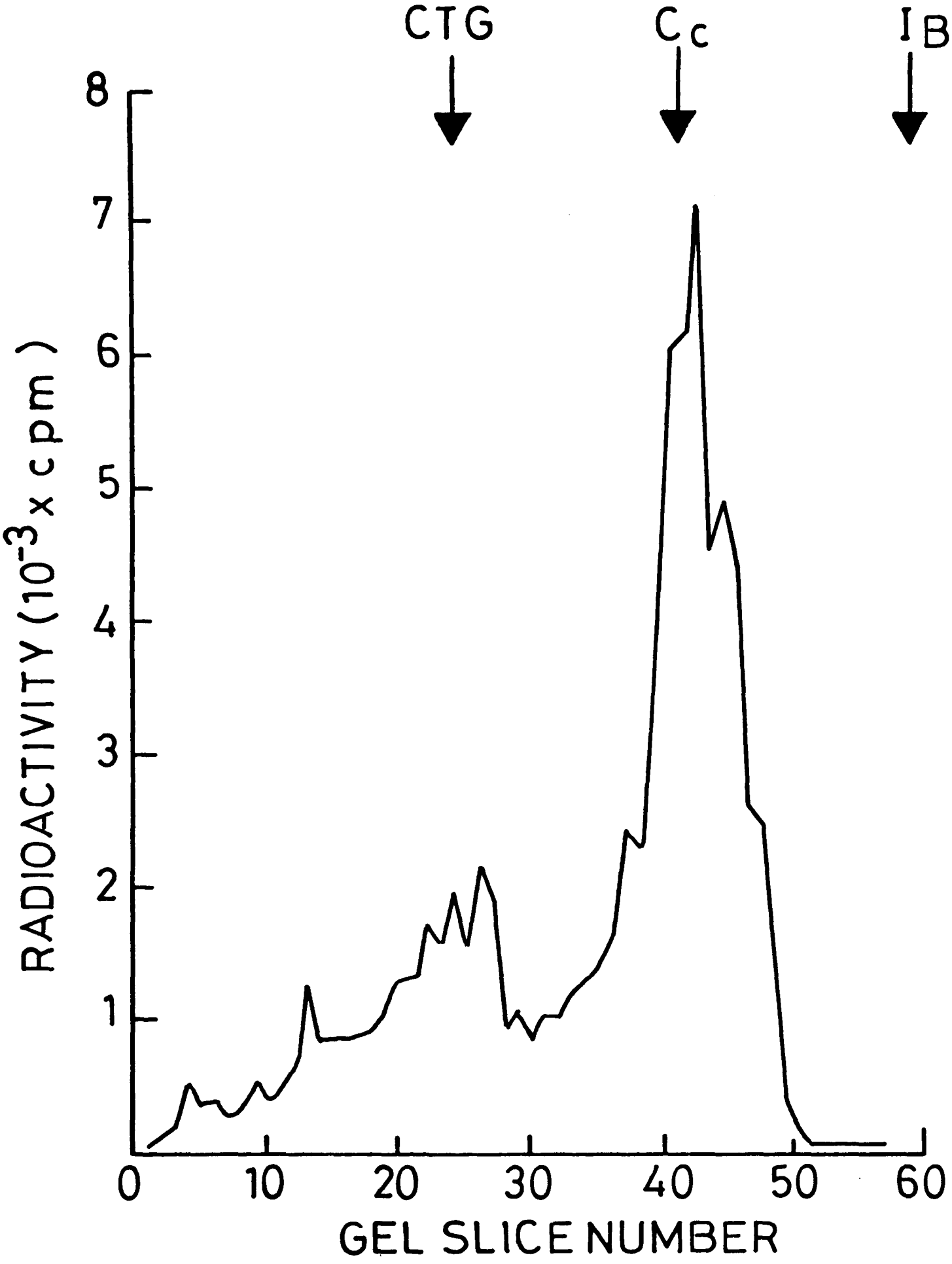
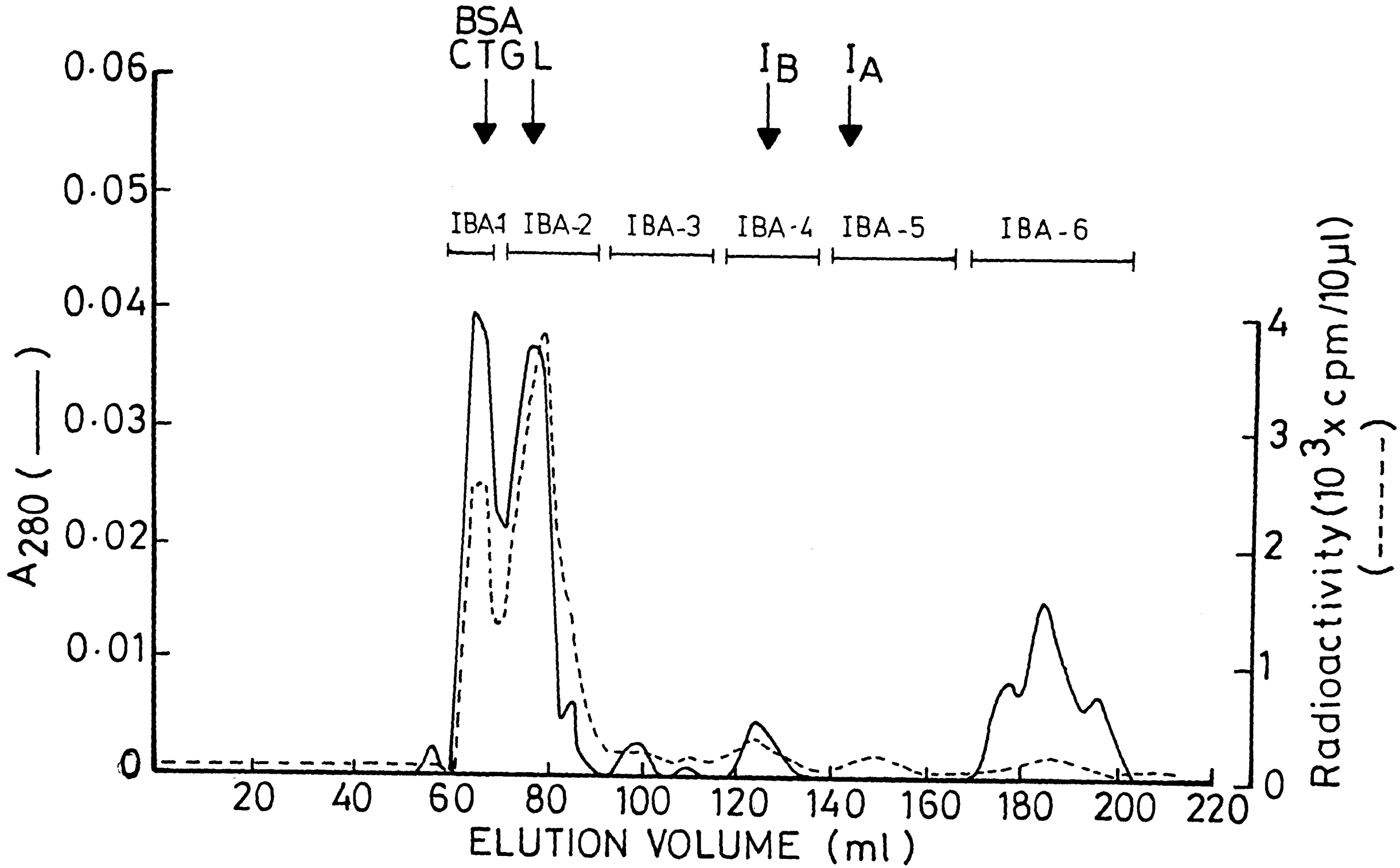


Fig. 4.10 SDS/polyacrylamide gel electrophoresis of factor $\bar{D}$ after cleavage with o-iodosobenzoic acid

The o-iodosobenzoic acid digest of ^{14}C-labelled factor $\bar{D}$ was purified on SDS/12% (w/v) polyacrylamide cylindrical gels. Gel slices were counted by liquid scintillation counting. Details are given in the text. Reduced and alkylated marker proteins were chymotrypsinogen (CTG, 25,700 mol. wt.), cytochrome c (C_c, 12,400) and insulin B chain (I_B, 3,400).

Fig. 4.11 Purification on Sephadex G-50 of peptides obtained by IBA digestion of factor $\bar{D}$

Peptides were eluted in 5% (v/v) formic acid. Details are given in section 2.2.11.2. Reduced and alkylated marker proteins were used to calibrate the column: bovine serum albumin (BSA, mol. wt. 67,000), chymotrypsinogen (CTG, mol. wt. 25,700), lysozyme (L, mol. wt. 14,300), insulin B chain (I_B, mol. wt. 3,400) and insulin A chain (I_A, mol. wt. 2,400). Column fractions were pooled as indicated by bar-lines.


BSA
CTGL
IB
IA
IBA-1
IBA-2
IBA-3
IBA-4
IBA-5
IBA-6
0·06
0·05
0·04
0·03
0·02
0·01
0
A280 (——)
20
40
60
80
100
120
140
160
180
200
220
ELUTION VOLUME (ml)
4
3
2
1
0
Radioactivity (10³ x cpm/10µl)
(- - - -)

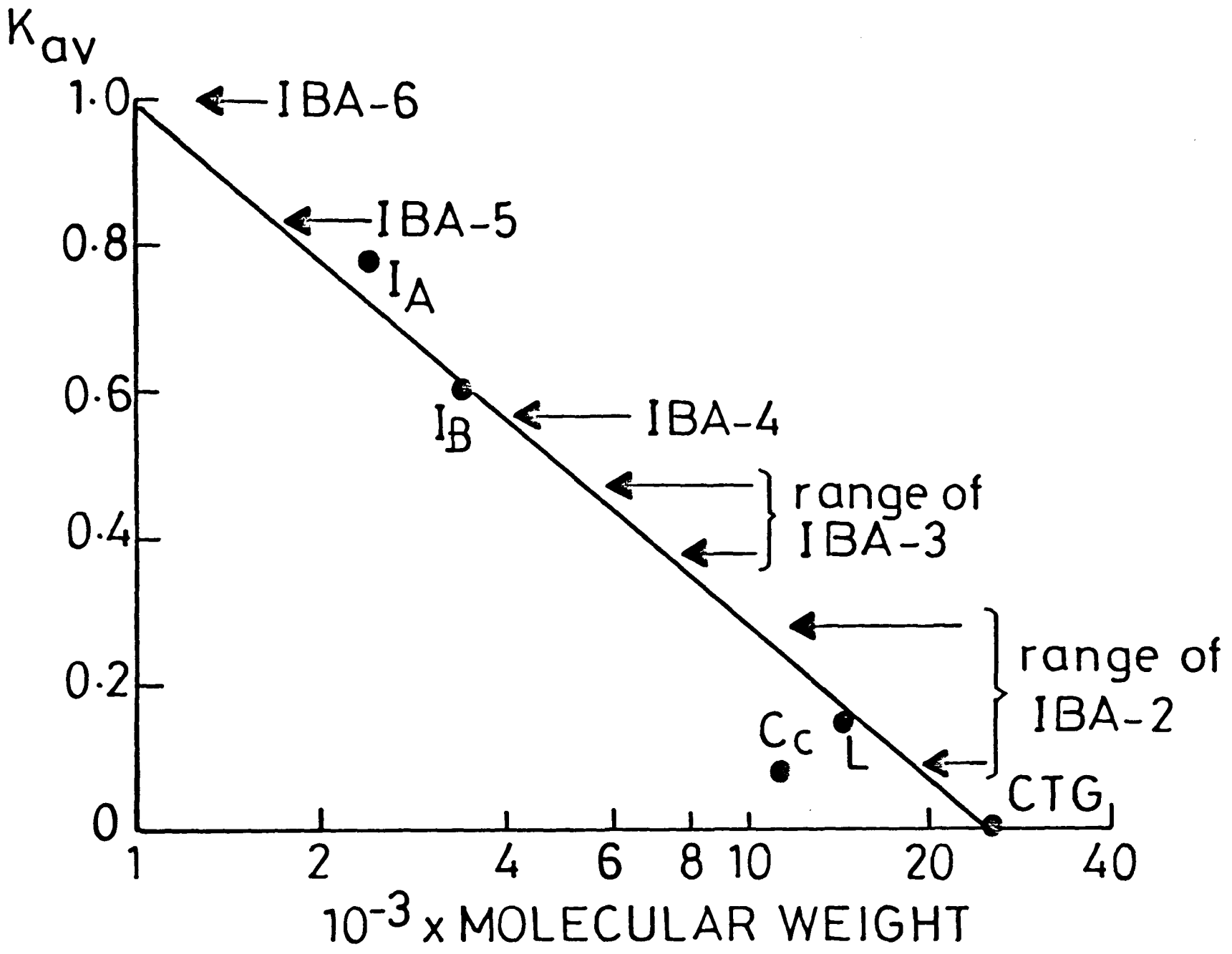


Fig. 4.12 Molecular weight estimation by gel filtration of IBA peptide pools

Data is taken from Fig. 4.11 and K_{av} was calculated as described in section 2.2.1.1. Molecular weights of marker proteins are: chymotrypsinogen (CTG, 25,700), lysozyme (L, 14,300), cytochrome c (C_c, 12,400), insulin B chain (I_B, 3,400) and insulin A chain (I_A, 2,400).

indicated only a slight quantitative difference in the composition of the two sides of the peak, the overall scan appearing similar to that of the pilot study (Fig. 4.10). To prevent further loss of material, the single protein peak obtained from Sephadex G-75 superfine was divided into three subfractions taken from the ascending, central and descending sections of the peak (IBA-2a, IBA-2b and IBA-2c respectively), and were sequenced without further purification. Three other low molecular weight fractions were identified from the Sephadex G-50 superfine column, fractions IBA-4 (mol. wt. 4,000), IBA-5 (mol. wt. 1,700) and IBA-6 (mol. wt. ≤500). The fraction IBA-3 (mol. wt. range 5,500-7,400) was also retained.

4.4.2.3 26 nmole digest

Factor $\bar{D}$ (26 nmoles) together with fraction IBA-1 obtained from the 70 nmole digest, was incubated with IBA and then purified on Sephadex G-50 superfine as described above. The elution profile was essentially the same as for the 70 nmole digest, except that fraction IBA-2 was enriched in the subfraction, IBA-2c. This enrichment was probably caused by further cleavage of the partially cleaved factor $\bar{D}$ contained in the fraction IBA-1 which was added from the 70 nmole digest. The column effluent was divided into fractions as for the 70 nmole digest, except that fraction IBA-2 was subdivided without prior repurification on Sephadex G-75 superfine.

After sampling for amino acid analysis, the fractions IBA-4, IBA-5 and IBA-6 were combined with their respective fractions from the 70 nmole digest. The combined fractions,

IBA-6, which contained the salt peak, were desalted twice on a column of Sephadex G-10.

4.4.3 Automated amino acid sequence analysis

The fractions obtained from Sephadex G-50 superfine, and in particular fraction IBA-2, frequently contained partial cleavage products and more than one peptide. The major peptides have therefore been named in their supposed order of alignment from the N-terminal of factor $\bar{D}$ as IBA-I, IBA-II, IBA-III and IBA-IV. Since the alignment of some of these peptides was only ascertained after comparison with tryptic peptides (section 4.5), reference to these has been necessary, and the relevant data is summarised in Fig. 4.33.

4.4.3.1 Peptides from fraction IBA-2

Since it was expected that the subfractions, IBA-2a, IBA-2b and IBA-2c would contain more than one peptide, amino acid analysis was not carried out, but the fractions were sequenced directly without further purification. Three major sequences were obtained (IBA-I, IBA-II and IBA-IV in Table 4.4 and Figs. 4.13-4.15). However, from the wide molecular weight range of fraction IBA-2 (10,000-19,000) it seemed likely that these may be partial cleavage products, containing 89-138 amino acids.

(1) Peptide IBA-I was found in subfractions IBA-2a and IBA-2b. This peptide had N-terminal isoleucine (total yield at cyle 1 from both subfractions was 8.8 nmoles) and the sequence was identical to that of the N-terminal of intact factor $\bar{D}$. The yields of PTH amino acids given in Table 4.4 and Fig. 4.13 were obtained from subfraction IBA-2b.

Table 4.4 Automated sequence analysis of major o-iodosobenzoic acid fragments of factor D̄

Residue Number	Amino acid identified in IBA-I	IBA-II	IBA-III[1]	IBA-IV
1	Ile (7.8)	Val (2.4)	Cly (5.2)	Gly (6.1)
2	Leu (7.7)	Leu (2.3)	-	Ile (7.1)
3	Gly (4.9)	Thr (0.4)	Val (2.5)	Val (7.3)
4	Gly (4.9)	Ala (1.4)	Asp (1.1)	Asn (4.2)
5	Arg (0.6)	Ala (1.4)	-	His (0.7)
6	Glu (4.3)	-	Asp (0.8)	Ala (6.3)
7	Ala (3.6)	Cys	Val (0.4)	Gly (3.4)
8	Glu (3.7)	Leu (1.0)	Ala (0.6)	Arg (1.3)
9	Ala (2.9)	Glu (0.6)	Pro (0.1)	Arg (1.7)
10	His (0.4)	Asp (0.4)	Gly (0.2)	Pro (2.7)
11	Ala (2.3)	Ala (0.8)	Thr (0.1)	Asp (1.4)
12	Arg (0.7)	Ala (0.6)	Leu (0.2)	-
13	Pro (0.8)	(Gly) (0.4)	(Cys)	Leu (3.4)
14	Tyr (0.9)	Gly (0.5)	-	Gln (2.0)
15	Met (1.7)	(Tyr) (0.3)	Val (0.3)	His (0.3)
16	Ala (1.7)	Val (1.1)	Ala (0.1)	Val (2.4)
17	Ser (0.4)	-		Leu (1.9)
18	Val (2.3)	Val (0.8)		Leu (2.3)
19	Gln (0.8)	Leu (0.8)		Pro (0.8)
20	Leu (1.9)	Leu (0.7)		Val (1.6)
21		Gly (0.3)		Leu (2.1)
22		Ala (0.3)		Asp (0.8)
23		(Ile) (0.6)		-
24				Ala (1.0)
25				Lys (2.4)
26				(Cys)
Amount[2] used in sequencer	25	19	42	31

Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields (in nmoles) are given in brackets. The presence of SCM-cysteine PTH derivatives was confirmed by detection of radioactive label.

Residues in brackets are only tentatively identified.

-, not identified.

[1] his and arg not determined

[2] quantities used in the sequencer were calculated theoretically, as fractions were not pure (section 4.4.3.5).

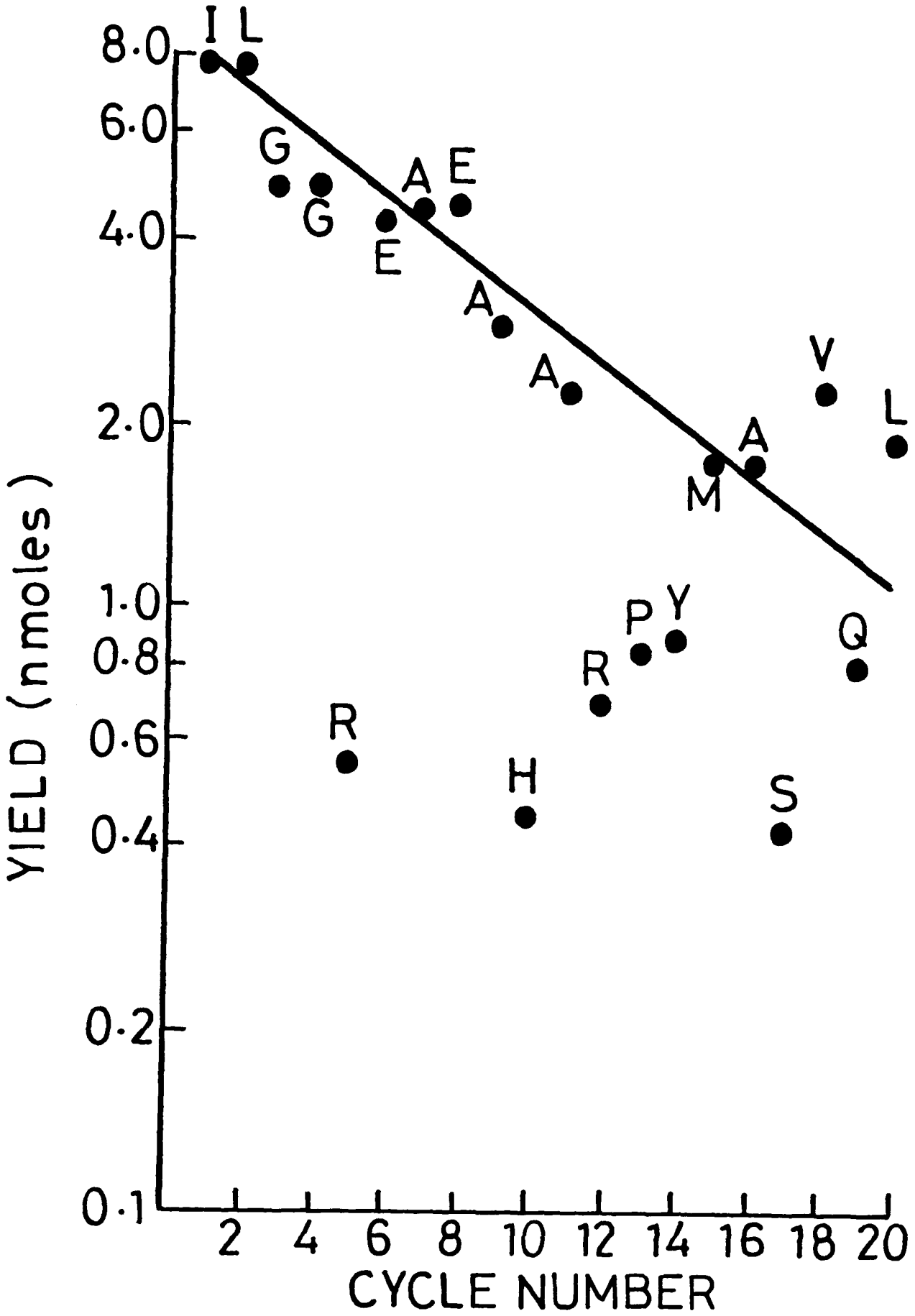


Fig. 4.13 Yields per cycle for automated sequence analysis of peptide IBA-I

The amino acid sequence, denoted by the one letter code, is also found in Table 4.4. The stepwise yield, calculated from ala, 7,9,11 and 16, was 92%.

(2) Peptide IBA-IV (Fig. 4.14) was found in fraction IBA-2b and was present in a similar yield to peptide IBA-I. The peptide, which had N-terminal glycine, was homologous in positions 1-8 with the tryptic peptide, T6-II residues 14-21, and in positions 9-22 with the tryptic peptide T6-III, residues 1-14. Of the residues which were not obtained in the sequence, residue 12 was identified as serine and residue 23 as arginine, by analysis of the tryptic peptides. By comparison with other serine proteases (Dayhoff, 1979) it can be seen that in two regions of the molecule a -trp-gly- sequence may be expected. These two regions correspond to $-trp_{141}-gly_{142}-$, and $-trp_{215}-gly_{216}-$, in chymotrypsinogen. By comparison of conserved residues in the region of gly_{142} onward and gly_{216} onward with peptide IBA-IV, it was concluded that this peptide probably originated from the region corresponding to gly_{142} onward. Residues 1, 13, 19 and 26 could be aligned with the invariant residues gly_{142}, leu_{155}, pro_{161} and cys_{168} (chymotrypsinogen numbering) although identification of cys_{26} in peptide IBA-IV was not unequivocal. Other conserved residues which also occur in peptide IBA-IV are chymotrypsinogen residues 156 and 160 (corresponding to glu_{14} and leu_{18} of peptide IBA-IV).

(3) Peptide IBA-II was the major sequence of fraction IBA-2c (Fig. 4.15) and showed strong homology at the N-terminal end with the highly conserved sequence which surrounds the active site histidine of other serine proteases (his_{57} in chymotrypsinogen). In the sequence

-trp-val-leu-thr-ala-ala-his-cys-

valine and the tetrapeptide, -ala-ala-his-cys-, are invariant

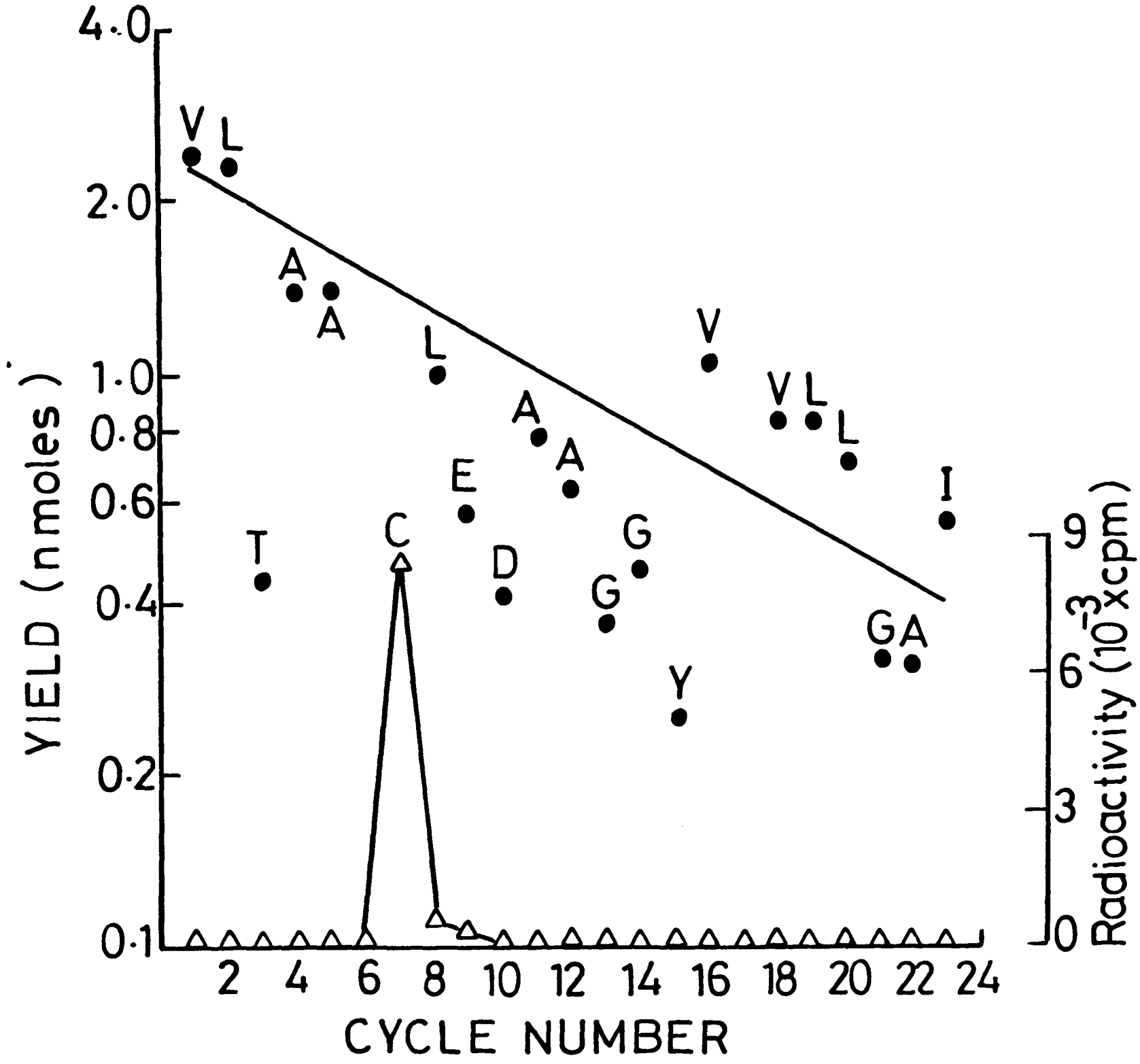


Fig. 4.15 Yield per cycle for automated sequence analysis of peptide IBA-II

The amino acid sequence denoted by the one letter code is also found in Table 4.4. Radioactivity released (Δ) is from ^{14}C-SCM-cysteine. The stepwise yield, calculated from leu 2,8,19 and 20 was 93%.

while the remaining residues are highly conserved. Unfortunately, residue 6, which by homology should be the active site histidine, was not obtained, probably because of its being modified by o-iodosobenzoic acid. The data is similar from residues 1-22 to that obtained by Volanakis et al (1980) using intact factor $\bar{D}$ (residues 36-57) except for residues 3, 6-9, 13 and 14 which were unidentified by Volanakis et al, and residues 15 and 17. Residue 15 was identified as tyrosine in this study and as lysine by both Volanakis et al and by Davis (1980). The identification as tyrosine in this study was not unequivocal and requires confirmation. Residue 17 was not obtained, but Volanakis et al have identified glutamine at this point. Since a minor sequence at approximately one-fifth the concentration of the major sequence was also observed in the sequencer run on fraction IBA-2c, the data reported here require confirmation. However, the similarity with the data of Volanakis et al and the homology with the active site of other serine proteases both support the alignment of the peptide as suggested here.

4.4.3.2 Fraction IBA-4

The amino acid composition of fraction IBA-4 is found in Table 4.5. Although not identical, the composition is very similar to that of the N-terminal 35 residues of factor $\bar{D}$. Isoleucine was the N-terminal amino acid. This, together with the apparent molecular weight (4,000) and the similarity of amino acid composition, suggested that the major peptide of this fraction was residues 1-35 of factor $\bar{D}$. Automated sequence analysis was not carried out.

Table 4.5 Amino acid compositions of IBA fractions IBA-4 and IBA-5

Amino Acid	Amino acid composition (mol/mol) of fraction	
	IBA-4	IBA-5
Asp	1.9 (2)	1.9 (2)
Thr	0.7 (1)	0.7 (1)
Ser	1.5 (1)	1.1 (1)
Glu	4.6 (5)	2.0 (2)
Pro	1.4 (1)	0.7 (1)
Gly	5.1 (5)	3.1 (3)
Ala	4.6 (5)	1.7 (2)
Val	3.1 (3)	1.9 (2)
Ile	1.0 (1)	0.7 (1)
Leu	3.5 (4)	1.2 (1)
Tyr	0.8 (1)	-
Phe	-	-
His	1.8 (2)	0.4 (0)
Lys	0.8 (1)	0.5 (0)
Arg	1.6 (2)	0.8 (1)

The amino acid compositions are given as mol of residue/mol of peptide. The samples were hydrolysed at 110^{o}C for 48 h. No correction was made for serine and threonine destruction. Tryptophan, cystine and methionine were not determined. -, less than 0.3 mol of residue/mol of peptide.

4.4.3.3 Fraction IBA-5

Fraction IBA-5 contained one major sequence, IBA-III, with N-terminal glycine (Fig. 4.16). The background PTH amino acids were significant at the start of the run, but it was possible to identify a major sequence by comparison with tryptic peptides. The sequence overlapped with the tryptic peptide T6-II residues 1-11, except that in the IBA peptide, residues 5 and 14 (later recognised as arginine and aspartic acid respectively, section 4.5) were not identified. The alignment of the peptide was deduced following alignment of peptide T6-II and is also discussed in section 4.5. However, the IBA peptide shows homology at residues 10,11 and 13 with the conserved residues, gly_{133}, thr_{134} and cys_{136} (chymotrypsinogen numbering) which are found in other serine proteases, although the identification of half-cystine at position 13 of the peptide was only tentative.

The background of the sequencer run on fraction IBA-5 contained a sequence which began, ile-asp- (yield of isoleucine at cycle 1 = 1.3 nmoles). A short peptide containing the sequence

-trp-ile-asp-

was also identified by sequence analysis of tryptic peptide T7a-II (section 4.5). The impurities which were identified in the sequencer run on fraction IBA-5 were also reflected in the amino acid composition (Table 4.5) which was similar but not identical to the composition expected from the sequence data.

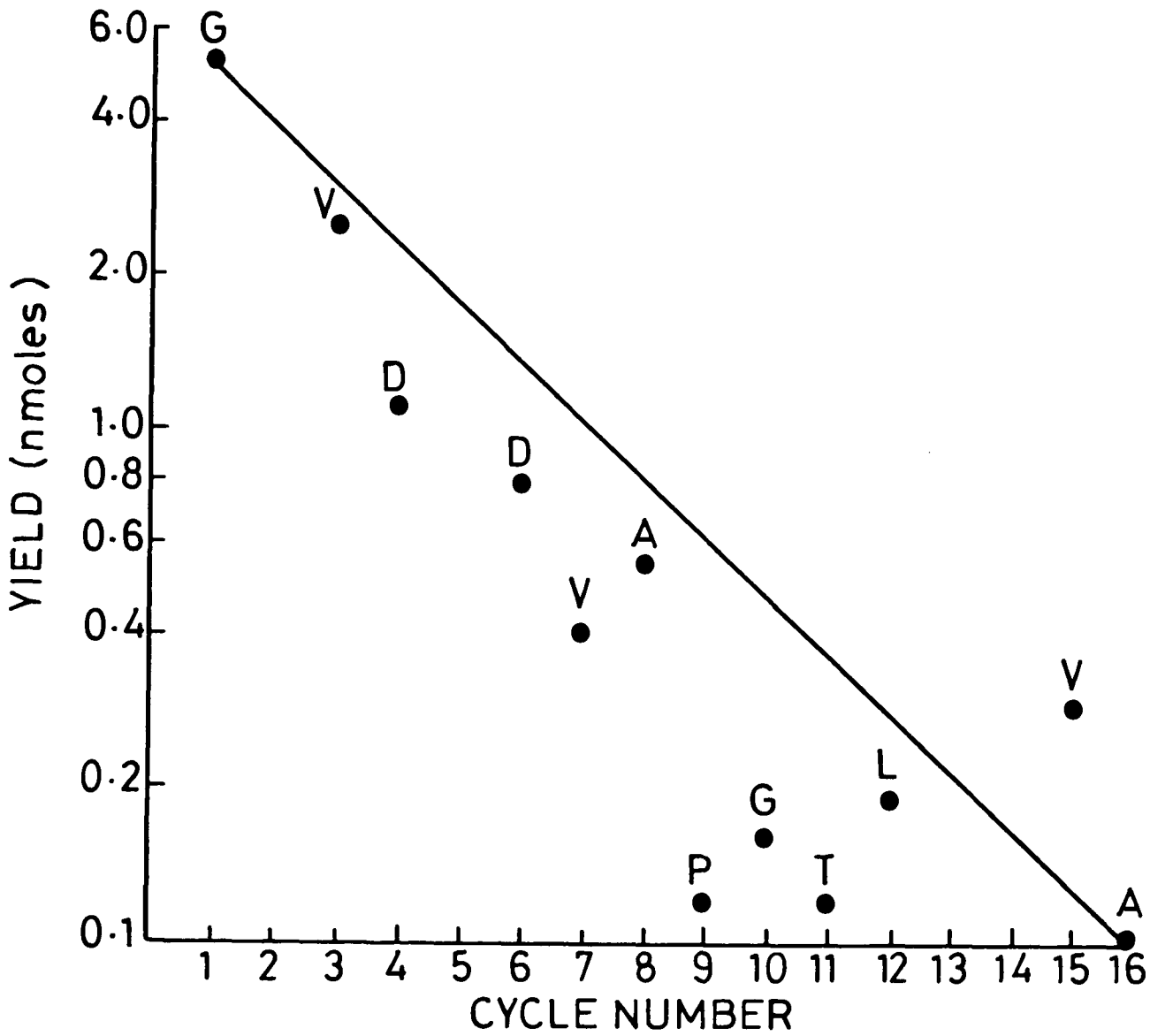


Fig. 4.16 Yield per cycle for automated sequence analysis of peptide IBA-III

Amino acids are denoted by the one letter code, and the sequence is found also in Table 4.4. The stepwise yield, calculated from val 3, 7 and 15 was 75%.

4.4.3.4 Fractions IBA-3 and IBA-6

No useful data were obtained from these fractions. Although the chromatography for N-terminal amino acids on subfractions from IBA-6 suggested the presence of a peptide with N-terminal isoleucine, attempts to purify this by Sephadex G-10, HPLC and pH 1·9 electrophoresis were all unsuccessful. An amino acid sequencer run also gave no information.

4.4.3.5 Discussion

From the sequence analysis of the IBA peptides of factor $\bar{D}$, four major amino acid sequences were identified. The alignment of these peptides is given in Fig. 4.17.

Peptide IBA-I was identified by its N-terminal amino acid, amino acid composition and molecular weight as the N-terminal 35 residues of factor $\bar{D}$. The peptide was isolated from fraction IBA-4 but was also present in subfractions IBA-2a and IBA-2b. However, from the apparent molecular weight of these subfractions (10,000-19,000) and the yield of the peptide from fraction IBA-4 (approx. 36%) it was evident that cleavage at trp_{35} was not complete.

Peptide IBA-II had N-terminal valine (residue 36 of factor $\bar{D}$) and by homology with other serine proteases probably contains the active site histidine, although this residue itself (residue 41 of factor $\bar{D}$; residue 6 of peptide IBA-II) was not identified. The apparent molecular weight of the peptide isolated from subfraction IBA-2c was 10,000 (approx. 90 amino acids). Although the background of PTH amino acids was high, a similar sequence could be detected in

Fig. 4.17 Alignment of major IBA peptides

Peptides IBA-I and IBA-III have apparent mol. wts. estimated by gelfiltration of 4000 and 1700 and contain approx. 35 and 18 amino acid residues respectively. The mol. wts. of peptides IBA-II, IBA-IV and IBA-V were approximated as 8000, 9000 and 500 respectively and would therefore contain approx. 73, 81 and 4 amino acid residues respectively. ⟶ , ( ) or a blank space denote positive, tentative or no identification respectively. | denotes supposed cleavage sites by o-iodosobenzoic acid.

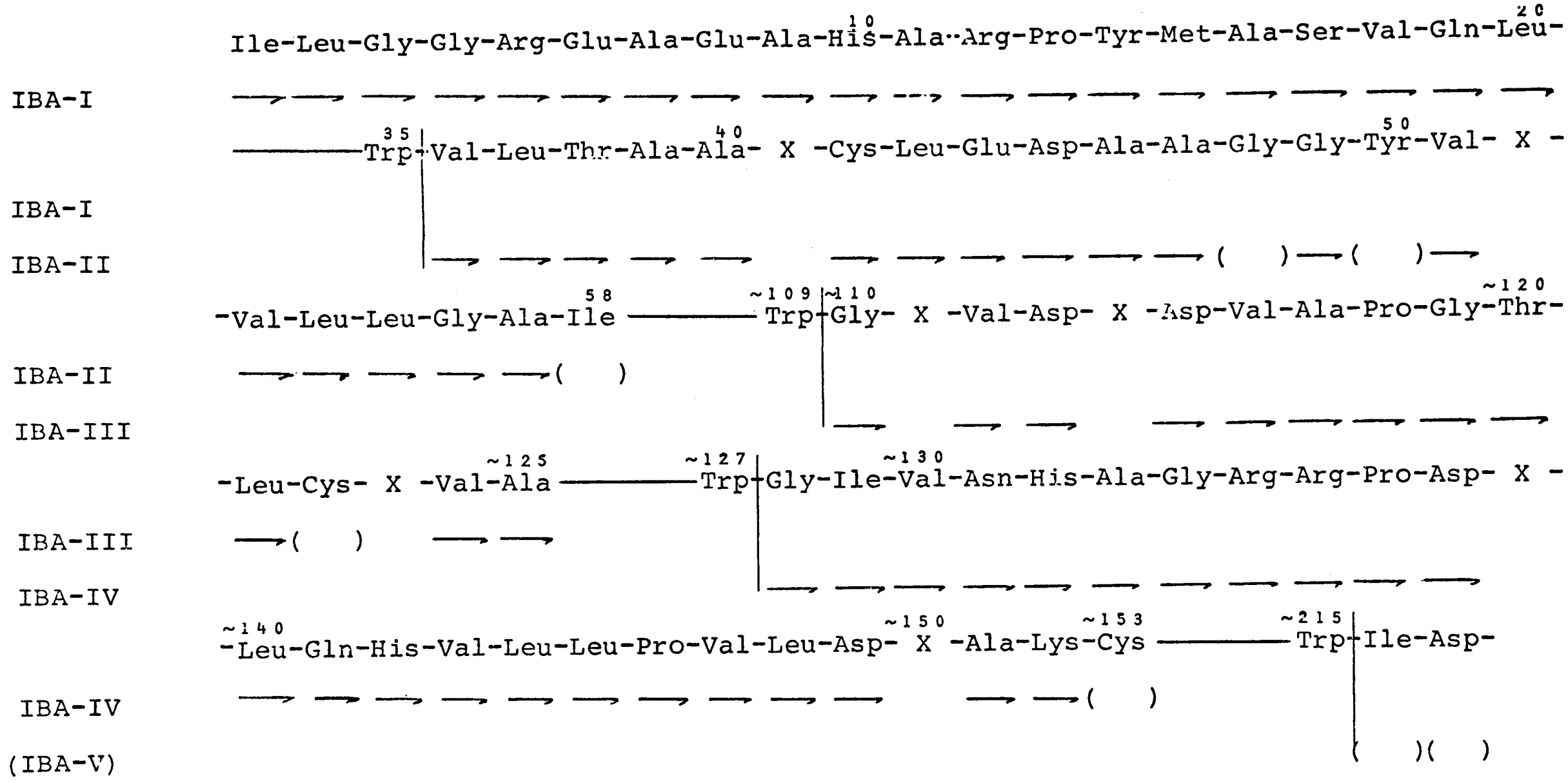


Fig. 4.17 Alignment of o-iodosobenzoic acid-derived peptides

the sequencer run on fraction IBA-3 (mol. wt. range 5,500-7,400) suggesting that the peptide isolated from fraction IBA-2c may be composed of both peptides IBA-II and IBA-III (cf. tryptic peptides, section 4.5).

Peptide IBA-III was identified in fraction IBA-5 and its alignment directly following peptide IBA-III was clarified by comparison with the tryptic peptide T6-II. The alignment of peptide IBA-IV which follows peptide IBA-III, was also elucidated by comparison with tryptic peptides T6-II and T6-III.

The possibility of a fifth very small peptide was suggested by (1) the finding of a short secondary sequence beginning ile-asp- in fraction IBA-5,

(2) the occurrence in tryptic peptide T7a-II of a similar sequence, -trp-ile-asp-,

(3) the tentative identification of an isoleucine N-terminal in the fraction IBA-6, although no sequence data was obtained on this fraction.

Although the occurrence of the secondary sequence in fraction IBA-5 would suggest a peptide of about 15 residues long, whereas a sequence in fraction IBA-6 would be <5 residues long, it is conceivable that the putative peptide could aggregate easily, particularly, if it were hydrophobic. However, the occurrence of five IBA peptides would suggest that the tryptophan analysis of factor $\bar{D}$ (2.7 residues/molecule) is incorrect, and that the true value is 4 residues/molecule. Further studies will be necessary to clarify this point.

Of the original factor $\bar{D}$, 4% remained uncleaved following digestion with IBA. However, the yields of the peptides varied and this probably reflected variation in the susceptibility of the trp-X bonds to cleavage. The yields of peptides IBA-I and IBA-III were 36% and 59% respectively. The yields of peptides IBA-II and IBA-IV were not easy to estimate as they were not completely purified. However, from their alignment (Fig. 4.17), an approximate yield can be calculated for cleavage at their respective trp-X bonds, allowing for the additional purification step that these peptides underwent. Assuming an 80% recovery of peptides from the gel filtration column, the yield of protein with peptide IBA-II as N-terminal would be 29% and for peptide IBA-IV, 47%. The estimated quantities of peptide used in the sequencer (Table 4.4) were calculated assuming these yields.

The apparent molecular weights of peptides IBA-I and IBA-III were 4,000 and 1,700 respectively. Because of the incomplete cleavage at tryptophanyl residues and the difficulty in separating the resultant IBA peptides, it was not possible to obtain estimates of the molecular weights of peptides IBA-II and IBA-IV. However, maximum values of 10,000 and 15,000 could be obtained from their elution positions on Sephadex G-50 (superfine). The sum of the apparent molecular weights of the four peptides (30,700) is then considerably more than the apparent molecular weight of factor $\bar{D}$ (24,000). It is probable, therefore, that these values are overestimates and that the peptides isolated from fraction IBA-2 were incomplete cleavage products.

4.5 CLEAVAGE AT ARGINYL RESIDUES USING TPCK-TREATED TRYPSIN

4.5.1 Introduction

Trypsin cleaves peptide bonds on the carboxyl side of arginine and lysine residues. Hydrolysis after these residues is usually decreased if the residue immediately following is proline. An acidic residue immediately preceding or following arginine or lysine will also diminish hydrolysis.

It is possible to modify the ε-amino group of lysine reversibly or irreversibly. Succinic anhydride modifies lysine residues irreversibly (Gounaris and Perlmann, 1967) and has the added advantage of incorporating a charged group into lysine so that the solubility of the peptide is maintained.

Factor $\bar{D}$ contains approximately 15 arginines and approximately 8 lysines per molecule. Tryptic digestion would therefore yield about 24 peptides. However, covalent modification of lysine residues will restrict the activity of trypsin to cleavage at arg-X, so that about 16 peptides would be obtained.

4.5.2 Digestion with trypsin and purification of peptides

Factor $\bar{D}$ (100 nmoles) was reduced and alkylated using an iodo-[2-^{14}C]-acetic acid label, then succinylated before digestion with trypsin for 4 h. After ultracentrifugation to remove insoluble material, the digest was initially fractionated on Sephadex G-75 superfine (Fig. 4.18). Nine major fractions were detected and were named T1 to T9 according to the order of elution. The fractions T4, T5 and

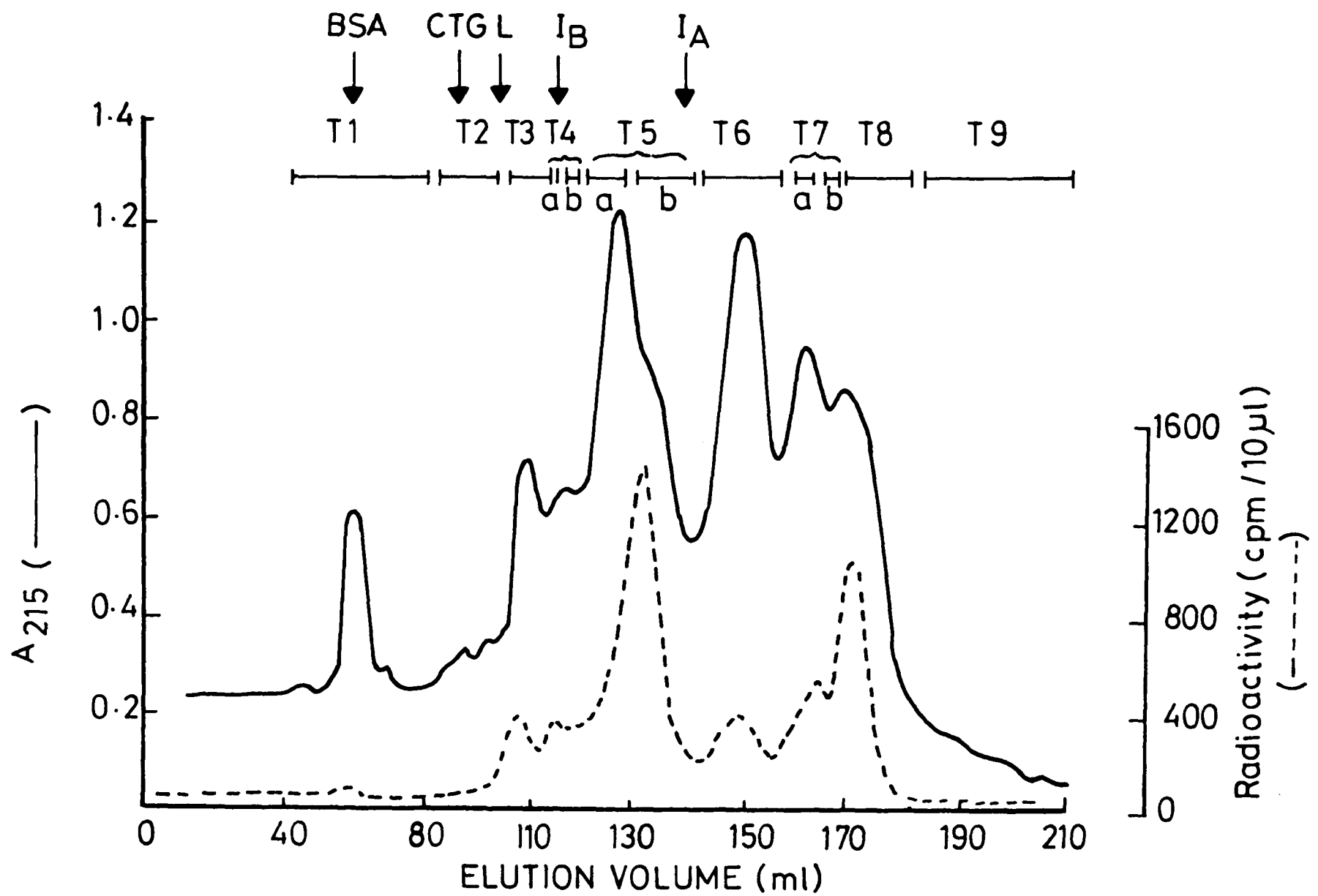


Fig. 4.18 Purification on Sephadex G-75 superfine of peptides obtained by digestion of succinylated factor D with trypsin-TPCK

Peptides were eluted in 0.1 M NH_4HCO_3. Details are given in section 2.2.11.3. Reduced and alkylated marker proteins were used to calibrate the column: bovine serum albumin (BSA, 67,000), chymotrypsinogen (CTG, 25,700), lysozyme (L, 14,300), insulin B chain (I_B, 3,400) and insulin A chain (I_A, 2,400). Column fractions were pooled as indicated by bar lines.

T7 were each divided further into subfractions, a and b, as in each case the peak of protein and radioactivity did not coincide. The N-terminal amino acids of fractions T5a, T5b and T6 were determined by dansylation, but in each case more than one DNS-amino acid was detected. Therefore, each of the fractions from T2 to T8 were further purified by HPLC. Major peaks obtained by HPLC were numbered according to their order of elution, and in the discussion which follows, this number follows the G-75 fraction number. The apparent molecular weights stated were estimated by gel filtration (Fig. 4.19).

4.5.3 Automated amino acid sequence analysis

4.5.3.1 Fractions T1, T2, T3, T4

Fraction T1 (average molecular weight 67,000) was not analysed. It most probably contained aggregated protein. Fraction T2 (average molecular weight, 23,000) and T3 (average molecular weight, 10,500) were relatively insoluble in the aqueous buffers used for HPLC and the apparent yields of peptides derived from these fractions were correspondingly low (less than 10 nmoles). Fraction T2 most likely contained uncleaved factor $\bar{D}$ and /or aggregated peptides. The total amino acid content of T2 was approximately 0.5% of the theoretical content of the total factor $\bar{D}$ used in the digest.

The amino acid compositions of peptides from fraction T3 suggested that at least some of these were either partial cleavage products or aggregations of smaller peptides. However, this fraction also contained two peptides, one with N-terminal isoleucine and the other with N-terminal proline (peptide T3). From the mol. wts. and amino acid compositions

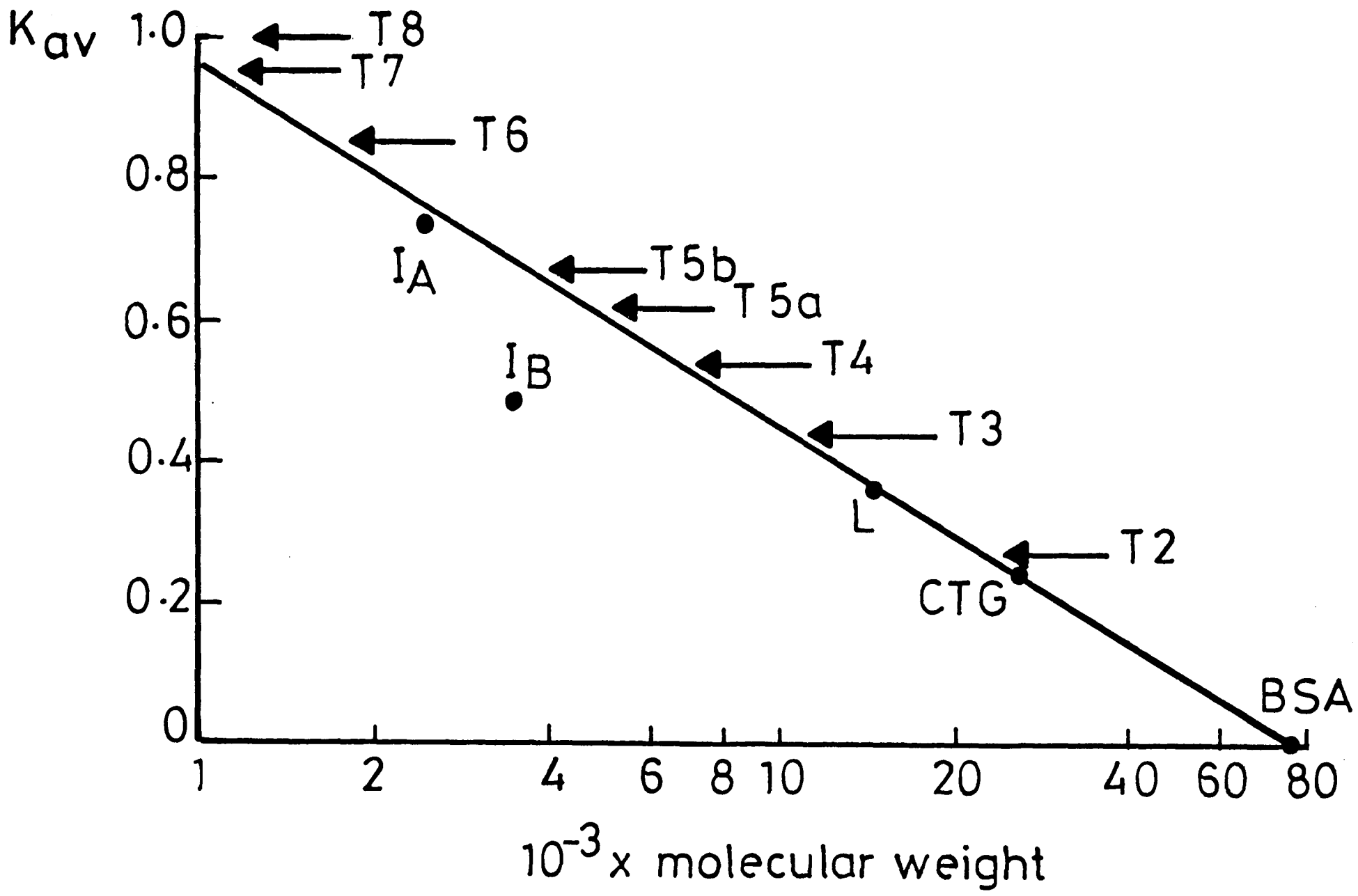


Fig. 4.19 Molecular weight estimation by gel filtration of tryptic peptide pools

Data is taken from Fig. 4.18. K_{av} was calculated as described in section 2.2.1.1. Molecular weights of marker proteins are: albumin (BSA, 67,000), chymotrypsinogen (CTG, 25,700), lysozyme (L, 14,300), insulin B chain (I_B, 3,400) and insulin A chain (I_A, 2,400).

of the peptides it seemed likely that their N-terminal amino acids corresponded to ile_1 and pro_{13} of intact factor $\bar{D}$. Peptide T3 had a similar but not identical amino acid composition to that of factor $\bar{D}$, residues 13-58 (Table 4.6). The apparent molecular weight of peptide T3 estimated by gel filtration, suggested that it would contain about 90 amino acids, although from the composition, a value of 60-70 amino acids seemed more likely. Sequence analysis was not carried out.

Fractions T4a and T4b were relatively soluble in aqueous buffers but the yields of peptides obtained by HPLC were less than 5 nmoles. Again, it appeared that these peptides were either partial cleavage products or derived by aggregation of smaller peptides.

4.5.3.2 Fractions T5a and T5b (3,800-4,800 mol. wt.)

(1) Fraction T5a (average molecular weight 4,800) contained one major peptide in a yield of 60 nmoles. Sequence data is given in Table 4.7 and Fig. 4.20. The N-terminal amino acid was alanine (yield 13.7 nmoles). The sequence obtained had not been found in earlier studies and was therefore aligned by homology with other serine proteases. From the molecular weight, a peptide of about 42 amino acids was expected. Residues 4, 5, 18, 19, 25, 31, 33 and 35 could be aligned with the conserved residues his_{91}, pro_{92}, leu_{105}, leu_{106}, ala_{112}, val_{118}, pro_{120} and pro_{124} (chymotrypsinogen numbering), and residues 15, 21 and 34 with asp_{102}, leu_{108} and leu_{123} which are invariant in other serine proteases. The sequence, $-leu_{34}-pro_{35}-trp_{36}-$, in peptide T5a was only tentatively

Table 4.6 Amino acid composition of peptide T3

Amino Acid	Amino acid composition (residues/100 residues)	
	T3	Factor $\bar{D}$ (residues 13-58)
Asp	7.9	4.3
Thr	2.7	2.2
Ser	8.2	2.2
Glu	11.6	10.9
Pro	4.6	2.2
Gly	11.3	13.1
Ala	12.3	17.4
Val	9.9	13.0
Cys	2.7	4.3
Met	1.1	2.2
Ile	1.7	2.2
Leu	12.5	15.2
Tyr	1.7	4.3
Phe	1.2	0
His	4.4	2.2
Lys	4.7	0
Arg	1.5	0
Other residues	N.D.	4.3

The amino acid composition of peptide T3 was calculated as residues/100 residues. The sample was hydrolysed at 110°C for 48 h. No correction was made for serine and threonine destruction. Cystine was measured as SCM-cysteine. N.D., not determined.

The composition of factor $\bar{D}$ (residues 13-58) was calculated from previous sequence data. "Other residues" include tryptophan and unidentified residues.

Table 4.7 Automated sequence analysis of tryptic peptides isolated from fraction T5

Residue number	Amino acid identified in T5a	Amino acid identified in T5b
1	Ala (13.7)	Asp (16.1)
2	Val (15.9)	Ser (0.7)
3	Pro (9.2)	Cys
4	His (16.6)	Lys (10.3)
5	Pro (7.3)	Gly (4.7)
6	Asp (13.2)	Asp (6.1)
7	Ser (0.7)	Ser (0.6)
8	Gln (6.2)	Gly (2.2)
9	Pro (2.7)	Gly (2.8)
10	Asp (6.9)	Pro (1.5)
11	Thr (0.5)	Leu (3.5)
12	Ile (3.4)	Val (4.5)
13	Asp (3.8)	Cys
14	His (2.6)	Gly (1.7)
15	Asp (3.1)	Gly (2.0)
16	Leu (3.6)	Val (2.9)
17	Leu (3.5)	Leu (1.0)
18	Leu (3.3)	Glu (1.3)
19	Leu (2.9)	Gly (1.1)
20	Gln (1.1)	Val (1.8)
21	Leu (2.2)	Val (1.8)
22	Ser (0.4)	(Thr)(0.1)
23	Glu (1.2)	-
24	Lys (0.5)	Gly (0.4)
25	Ala (0.9)	
26	Thr (0.1)	
27	Leu (0.9)	
28	Gly (0.4)	
29	Pro (0.3)	
30	Ala (0.5)	
31	Val (0.8)	
32	-	
33	Pro (0.3)	
34	(Leu)(0.6)	
35	(Pro)(0.2)	
36	(Trp)(0.02)	

Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields (in nmoles) are given in brackets. The presence of SCM-cysteine PTH derivatives was confirmed by detection of radioactive label. Lysine was detected as PTH-succinyl-lysine. Residues in brackets are only tentatively identified. -, not identified.

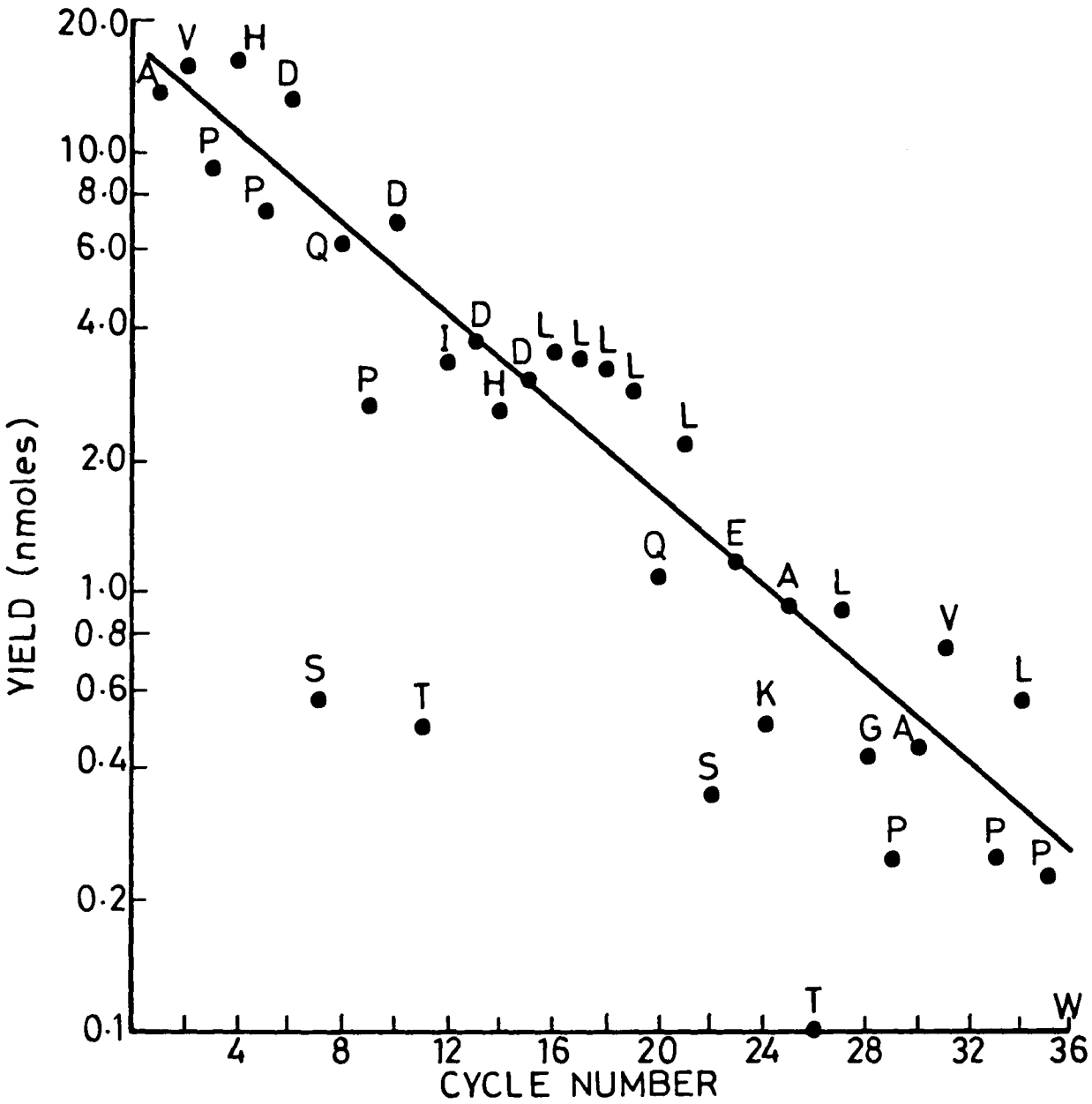


Fig. 4.20 Yield per cycle for automated sequence analysis of peptide T5a

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.7. The amount used in the sequencer was 62 nmoles. The stepwise yield calculated from leu 16, 17. 18, 19, 21 and 27 was 94%.

identified as at this point of the sequence, the background and low yields of PTH amino acids did not permit unequivocal identification. However, the conservation in serine proteases of leu and pro in positions equivalent to leu_{34}-pro_{35} supported the identity of the residues reported here. Additional evidence for trp_{36} was not obtained by comparison with other proteases, but was suggested by the observation that the N-terminal of IBA peptide IBA-III probably overlaps with the C-terminal of peptide T5a (see Fig. 4.33).

(2) Fraction T5b (average molecular weight 3,800) contained one major peptide which was recovered in a yield of 46 nmoles. It contained radioactivity (5.53×10^5 cpm). The N-terminal amino acid was asp and was obtained in a yield of 16.1 nmoles. The sequence (Table 4.7, Fig. 4.21) showed homology with the CNBr peptide, CNBr-2, in the first 19 amino acid residues and contained the putative specificity site residue (asp_1) and the active site serine (residue 7). The sequence was unambiguous for the first 21 residues. Thr_{22} was only tentatively assigned as it was obtained in very low yield.

The amino acid compositions of peptides T5a and T5b are given in Table 4.8. Comparison of sequence data with the amino acid compositions was not possible as neither sequence was obtained in its entirety. However, the similarity between the composition and sequence of peptide T5a suggested that most of the sequence of this peptide was obtained. An unexpected finding was the presence of two arginine residues in the amino acid composition of peptide T5b. Further sequence

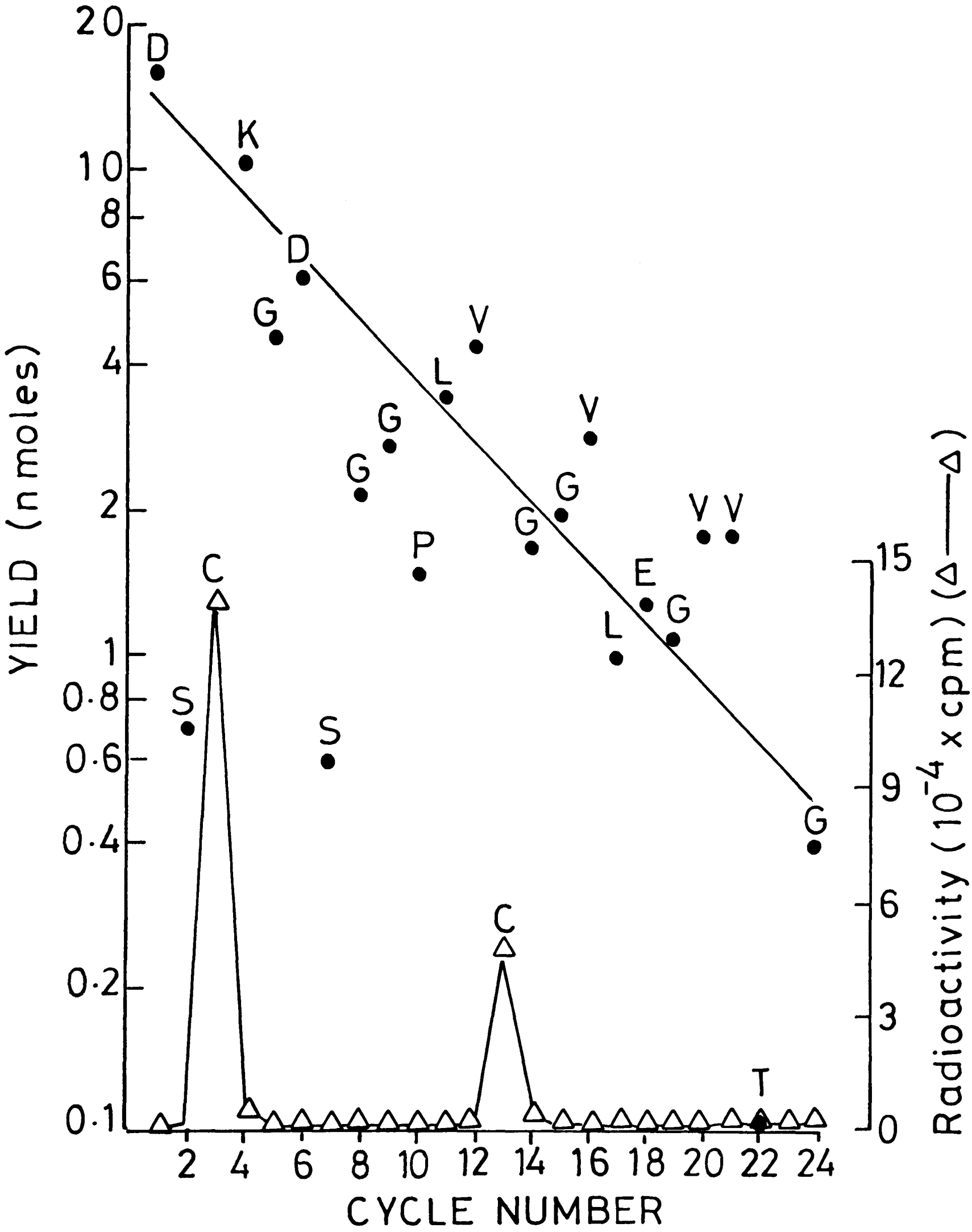


Fig. 4.21 Yield per cycle for automated sequence analysis of peptide T5b

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.7. Radioactivity released is from SCM-cysteine. The amount used in the sequencer was 46 nmoles. The stepwise yield, calculated from val 12, 16, 20 and 21 was 91%.

Table 4.8 Amino acid compositions of tryptic peptides from fraction T5

Amino acid	Amino acid composition (mol/mol) T5a	T5b
Asp	4.7 (5)	3.2 (3)
Thr	1.7 (2)	2.2 (2)
Ser	1.6 (2)	3.6 (4)
Glu	4.1 (4)	3.8 (4)
Pro	4.8 (5)	3.5 (3)
Gly	2.5 (2)	6.3 (6)
Ala	3.4 (3)	1.4 (1)
Val	2.6 (3)	4.3 (4)
Cys	-	1.5 (2)
Met	0.5 (0)	-
Ile	1.0 (1)	0.8 (1)
Leu	6.1 (6)	3.0 (3)
Tyr	-	0.5 (1)
Phe	-	-
His	1.8 (2)	0.4 (0)
Lys	1.3 (1)	1.7 (2)
Arg	1.4 (1)	2.0 (2)

The composition is given as mol of residue/mol of peptide. The samples were hydrolysed at 110^{o}C for 48 h. No correction was made for serine and threonine destruction and tryptophan was not determined. Cystine was estimated as SCM-cysteine. -, less than 0.3 mol of residue/mol of peptide.

data will be required before this observation can be clarified.

4.5.3.3 Fraction T6 (average mol. wt. 2,000)

Fraction T6 contained at least three peptides in yields greater than 10 nmoles. Only one of these (peptide T6-II) contained significant levels of radioactivity.

(1) Peptide T6-I (yield 52 nmoles) The sequence of this peptide (Table 4.9, Fig. 4.22) was obscured by artefacts present in the HPLC trace at several points, and will therefore require confirmation. Cycle one which in other sequencer runs was occasionally found to contain an artefact coinciding with the asp peak on HPLC, contained both asp and val in yields of 17.9 nmoles and 5.8 nmoles respectively. Cycles 2-6 were unambiguous, but cycle 7 contained a major residue (succinyl-lysine) as well as secondary residue (proline) which was identified as the major amino acid in cycle 8. This asynchronous pattern continued to the end of the run. It is possible that this ambiguity arose by partial cleavage of the succinyl-lysine during the coupling stage with PITC, a phenomenon which is known to occur more commonly with histidine. The amino acid composition calculated from the sequence was similar but not identical to the composition obtained by amino acid analysis (Table 4.10).

It has not yet been possible to align this peptide by homology with either other sequence data on factor $\bar{D}$ or other serine proteases. However, the sequence in serine proteases spanning CTG residues 72-84 is not highly conserved

Table 4.9 Automated sequence analysis of tryptic peptides isolated from fraction T6

Residue Number	Amino acid identified in T6-I	T6-II	T6-III
1	Asp (17.9)/Val (5.8)	Asp*	Arg(>40.0)
2	Leu (8.5)	Val (13.8)	Pro (7.5)
3	Ser (2.6)	Ala (11.1)	Asp (17.9)
4	Gln (6.4)	Pro (4.5)	Ser (1.8)
5	Pro (7.2)	Gly (10.8)	Leu (11.9)
6	Glu (6.4)	Thr (3.2)	Gln (8.9)
7	Lys (7.4)	Leu (7.4)	His (5.0)
8	Pro (2.5)	Cys	Val (5.6)
9	Lys (6.1)	Asp (9.8)	Leu (2.9)
10	Arg (1.5)	Val (5.8)	Leu (2.1)
11		Ala (3.4)	Pro (0.5)
12		Gly (3.1)	Val (1.8)
13		Trp (1.1)	Leu (1.1)
14		Gly (1.7)	Asp (0.9)
15		Ile (1.6)	Arg (0.6)
16		Val (1.9)	
17		Asn (0.6)	
18		His (0.4)	
19		Ala (0.6)	
20		Gly (0.5)	
21		Arg (0.4)	

Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields (in nmoles) are given in brackets. The presence of SCM-cysteine PTH derivatives was confirmed by detection of radioactive label. Lysine was detected as PTH-succinyl-lysine.

*The yield of asp in cycle 1 of peptide T6-II could not be calculated owing to an artefact in the HPLC trace which coincided with the asp peak. However, the presence of asp in cycle 1 was confirmed (1) by the finding of asp in cycle 4 of the secondary peptide which coeluted with peptide T6-II on HPLC and also overlapped with peptide T6-II in amino acid sequence, and (2) by the presence of DNS-asp as well as DNS-val in the determination of N-terminal amino acids of HPLC fraction T6-II by dansylation.

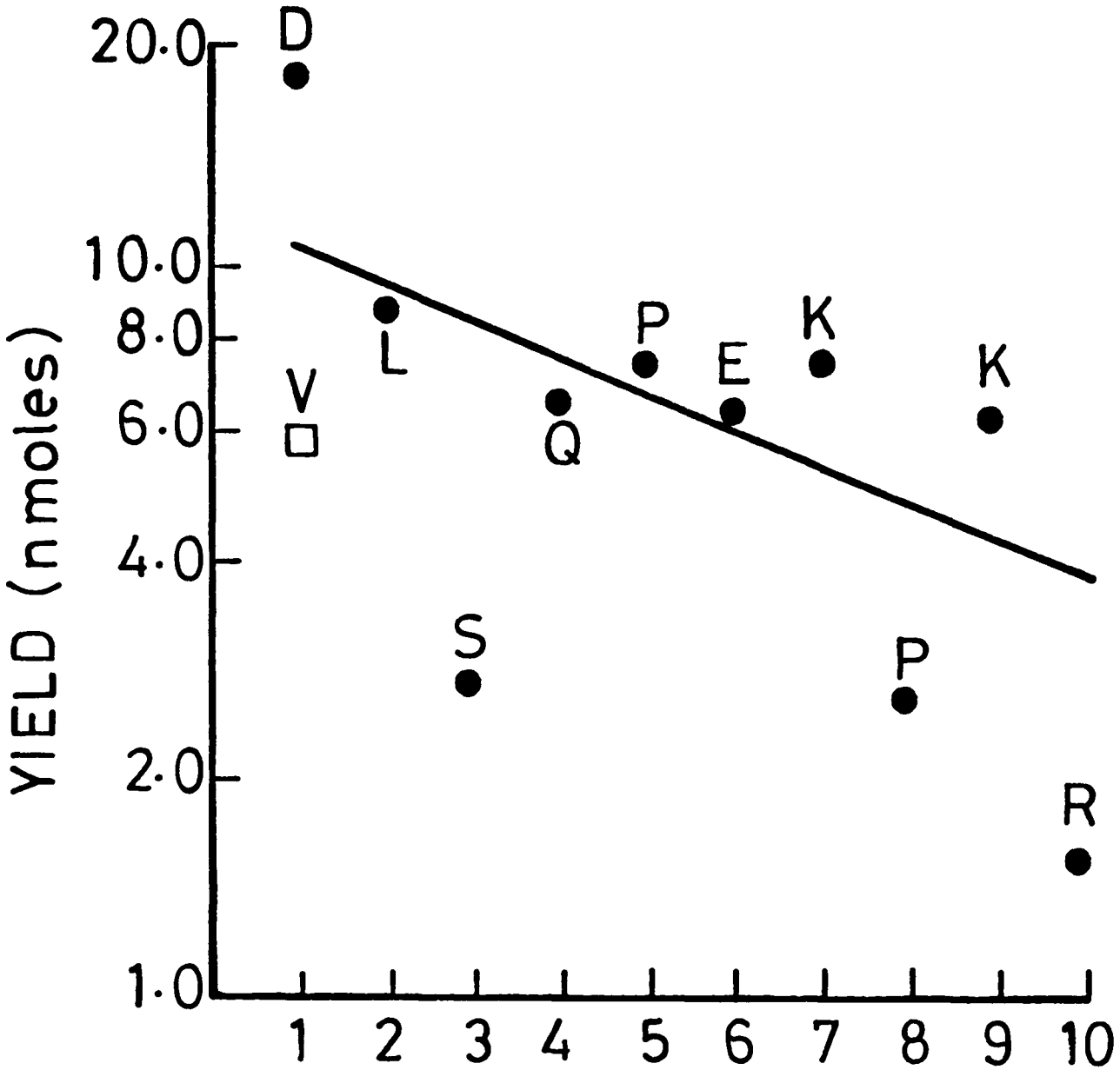


Fig. 4.22 Yield per cycle for automated sequence analysis of peptide T6-I

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.9. The amount used in the sequencer was 52 nmoles. The stepwise yield, calculated from lys 7 and 9 was 94%.

● major sequence, □ secondary residue at cycle 1.

Table 4.10 Amino acid compositions of tryptic peptides from fraction T6

Amino acid	Amino acid composition (mol/mol)		
	T6-I	T6-II	T6-III
Asp	1.2 (1)	3.1 (3)	2.2 (2)
Thr	-	0.8 (1)	-
Ser	1.0 (1)	0.6 (1)	0.8 (1)
Glu	1.7 (2)	0.9 (1)	1.2 (1)
Pro	3.1 (3)	1.8 (2)	2.0 (2)
Gly	1.3 (1)	3.9 (4)	0.5 (0)
Ala	-	2.8 (3)	-
Val	0.5 (0)	2.9 (3)	2.0 (2)
Cys	-	0.7 (1)	N.D.
Met	-	-	N.D.
Ile	-	1.0 (1)	-
Leu	0.6 (1)	1.6 (2)	3.8 (4)
Tyr	-	-	-
Phe	-	-	-
His	2.2 (2)	1.0 (1)	1.0 (1)
Lys	1.0 (1)	-	-
Arg	1.0 (1)	1.0 (1)	2.0 (2)

The amino acid compositions are given as mol of residue/mol of peptide. The samples were hydrolysed at 110°C for 48 h. No correction was made for serine and threonine destruction, and tryptophan was not determined. Cystine was measured as SCM-cysteine. N.D., not determined; -, less than 0.3 mol of residue/mol of peptide.

and no sequence data extending into this region has so far been obtained for factor $\bar{D}$. The region corresponds to the Ca^{2+}-binding site of trypsin and in most serine proteases contains at least three residues of glx. Until further evidence is available, peptide T6-I has been tentatively assigned to this region, but has not been included in the sequence of factor $\bar{D}$ given in Fig. 4.33. While it is possible that the sequence obtained may prove to be artefactual, the yield in which the peptide was recovered is consistent with its being a major cleavage product of factor $\bar{D}$.

(2) Peptide T6-II was recovered in a yield of approximately 31 nmoles and contained 320,000 cpm. The N-terminal amino acid was aspartic acid. The sequencer run was unambiguous (Table 4.9, Fig. 4.23) but contained a second peptide in about half the concentration of peptide T6-II. This peptide (data not shown) had the N-terminal sequence, val_1-asp_2-arg_3-, but was identical to peptide T6-II from residue 4 onwards. The incomplete cleavage at arg_3 in the secondary peptide was probably caused by the presence of asp at residues 2 and 4 as the occurrence of acidic residues immediately adjacent to arg decreases the rate of cleavage by trypsin at these points. The amino acid composition (Table 4.10) of fraction T6-II reflected the composition of the major peptide and the slight discrepancy in composition obtained by sequence analysis and amino acid analysis was not regarded as significant.

The sequence of peptide T6-II (residues 1-11) overlapped with that of IBA peptide IBA-III from residue 6 onwards (Table 4.4, Fig. 4.16) and the secondary peptide (residues 1-14) was

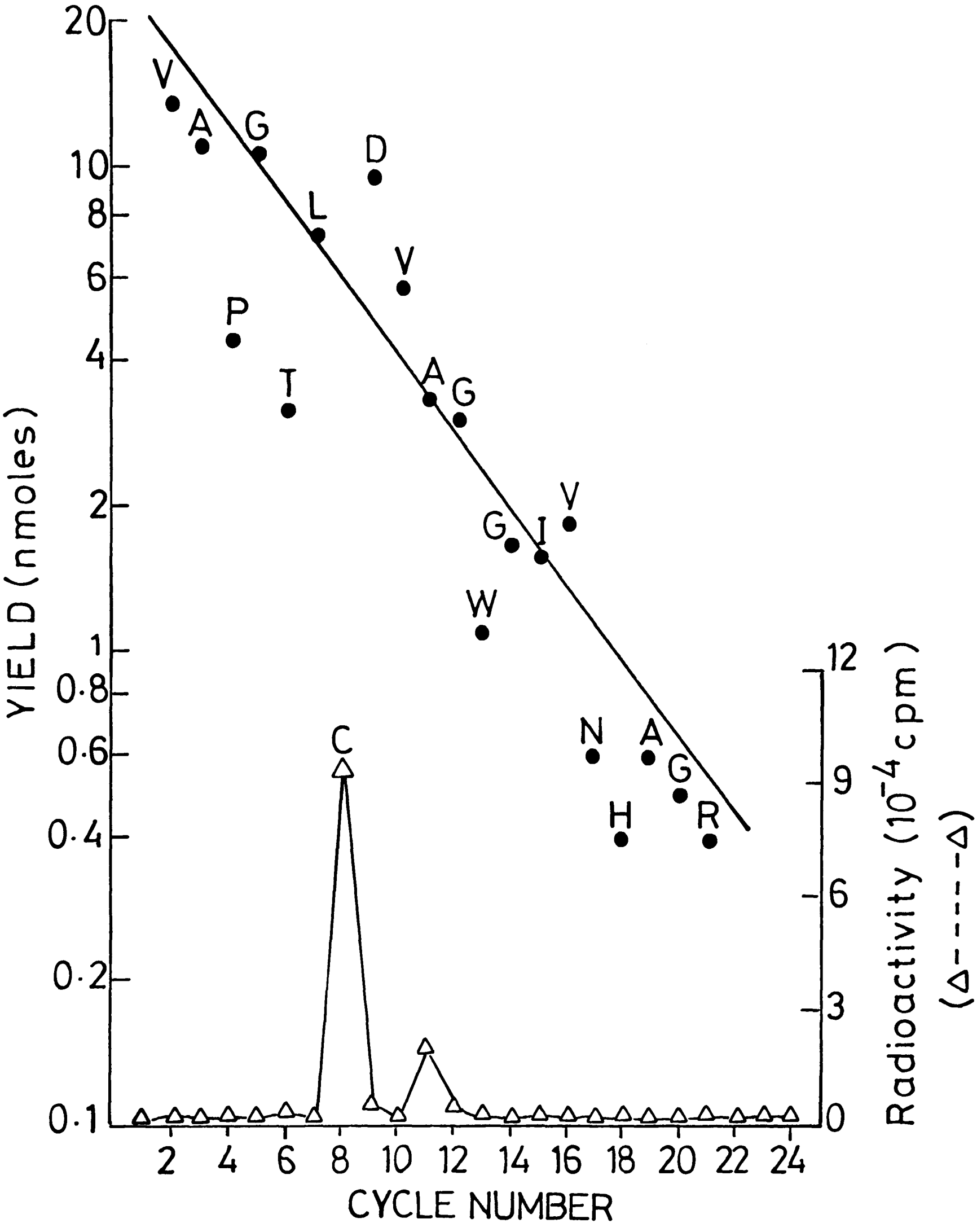


Fig. 4.23 Yield per cycle for automated sequence analysis of peptide T6-II

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.9. Radioactivity released (Δ) is from ^{14}C-SCM-cysteine. The amount used in the sequencer was approximated as 31 nmoles. The stepwise yield calculated from val 2, 10 and 16 was 90%.

identical to peptide IBA-III from residue 3 onwards. The C-terminal of peptide T6-II (residues 14-21) were identical to IBA peptide IBA-IV (Table 4.4, Fig. 4.14). This suggested that the sequence of peptide IBA-III reported in section 4.4.2.3 was lacking only the C-terminal two amino acids, -gly-trp.

By comparison with other serine proteases, peptide T6-II was placed after peptide T5a, as residues 12 and 14 could be aligned with the invariant residues gly_{140} and gly_{142} (CTG numbering) of other serine proteases, and residues 5,6,8 and 13 with the conserved residues gly_{133}, thr_{134}, cys_{136} and trp_{141}. The secondary peptide was also aligned in this position with residue 4 (asp) overlapping with residue 1 of peptide T6-II.

(3) Peptide T6-III (yield 50 nmoles) had N-terminal arginine (Table 4.9, Fig. 4.24). The sequence was unambiguous and the composition calculated from both the sequence and the amino acid analysis (Table 4.10) were identical. The N-terminal arginine suggests that the sequence around the site cleaved by trypsin was -arg-arg-pro-. Lack of cleavage at both sites of an -arg-arg- dipeptide is not unusual, and the preferred site of cleavage is probably explained by the presence of the sequence , arg-pro , which also does not cleave efficiently. The peptide was identical to IBA-peptide IBA-IV, residues 9-23. Thus by comparison with IBA peptides and by the conservation of leu_5 and pro_{11} (corresponding to invariant residues 155 and 161 in CTG respectively) and of gln_6 and leu_{10} (corresponding to conserved residues 156 and 160 respectively),

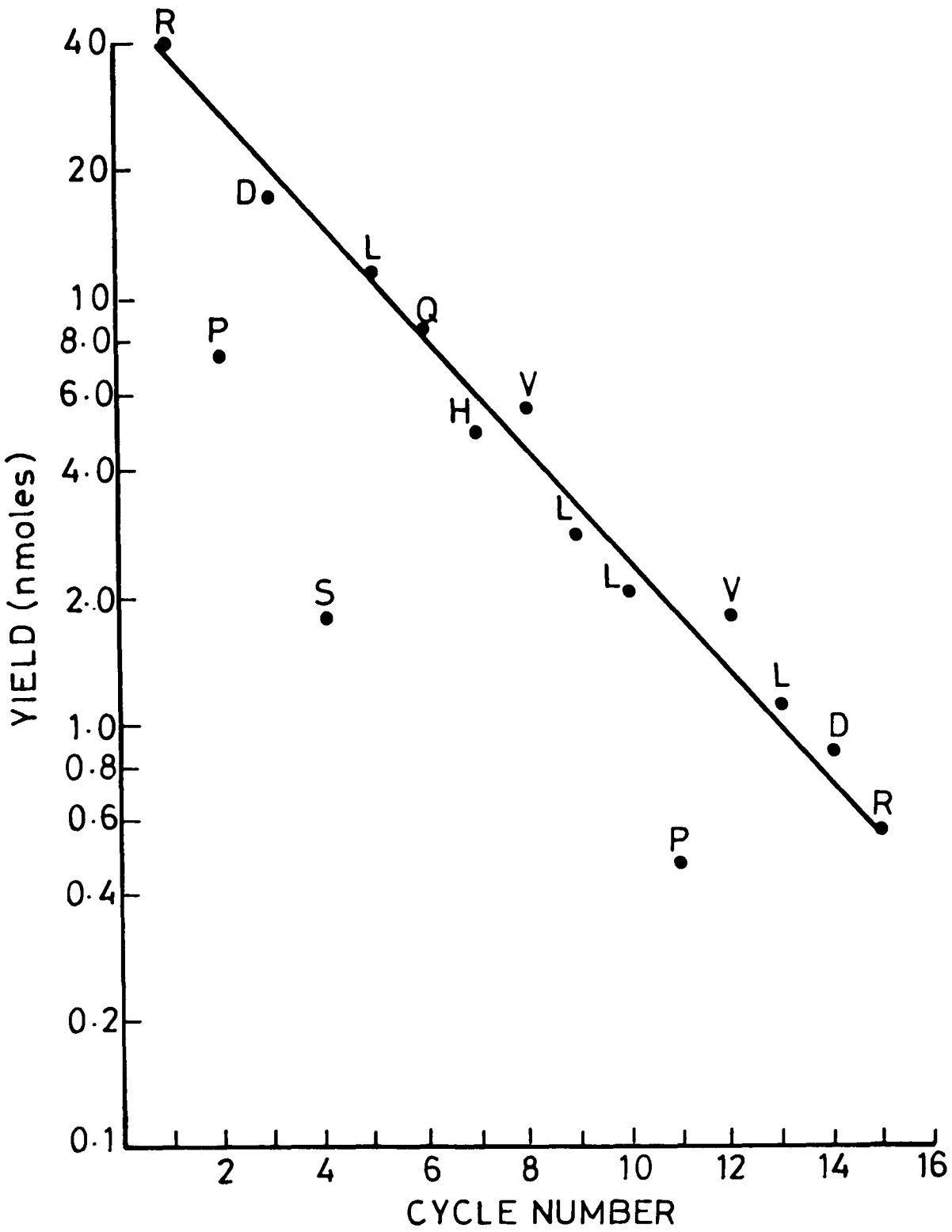


Fig. 4.24 Yield per cycle for automated sequence analysis of peptide T6-III

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.9. The amount used in the sequencer was 50 nmoles. The stepwise yield calculated from leu 5, 9, 10 and 13 was 81%.

peptide T6-III was aligned immediately adjacent to peptide T6-II.

4.5.3.4 Fraction T7 (mol. wt. range 600-1,300)

(1) Peptide T7a-I was obtained in an approximate yield of 20 nmoles. Although this yield was low, it also eluted in fractions T6 (19 nmoles) and T7b (4 nmoles) so that the total yield was approximately 43 nmoles. This unusual elution pattern was probably caused by its aggregation in the aqueous buffer used for gel filtration. The peptide contained 8 amino acids and the N-terminal residue was lysine (Table 4.11, Fig. 4.25).

The peptide was homologous with other serine proteases from residues 223 to 230 (CTG numbering) which has the highly conserved sequence

```
-arg-X-X-lys-pro-gly-val-tyr-thr-
222                          229
```

Although residue 230 is not highly conserved, it tends to be lys or arg. The sequence of a second peptide which coeluted with peptide T7a-I on HPLC, but was partially separated from it by gel filtration, was also obtained. This peptide was present at about one-quarter the concentration of peptide T7a-I, and was identical with peptide T7b-III. The amino acid composition of fraction T7a-I (Table 4.12) reflects the composition of both peptides.

(2) Peptide T7a-II was obtained in a low yield of approximately 10-15 nmoles. The amino acid composition (Table 4.12) contained no arginine, suggesting that this may be the

Table 4.11 Automated sequence analysis of tryptic peptides isolated from fraction T7

Residue number	Amino acid identified in		
	T7a-I	T7a-II	T7b-III
1	Lys (15.3)	Val (8.2)	Leu (5.1)
2	Lys (14.6)	Ala (3.9)	Met (4.7)
3	Pro (9.2)	Ser (0.5)	Cys
4	Gly (12.8)	Tyr (3.7)	Ala (3.4)
5	Ile (10.2)	Ala (2.6)	Glu (2.8)
6	Tyr (11.7)	Ala (1.7)	Ser (0.3)
7	Thr (1.0)	Trp (1.0)	Asn (1.2)
8	Arg (4.0)	Ile (1.3)	Arg (1.2)
9		Asp (1.0)	Arg (0.8)
10		(Val) (0.4)	

Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields (in nmoles) are given in brackets. The presence of SCM-cysteine PTH derivatives was confirmed by detection of radioactive label. Lysine was detected as PTH-succinyl-lysine. Residue in brackets, only tentatively identified.

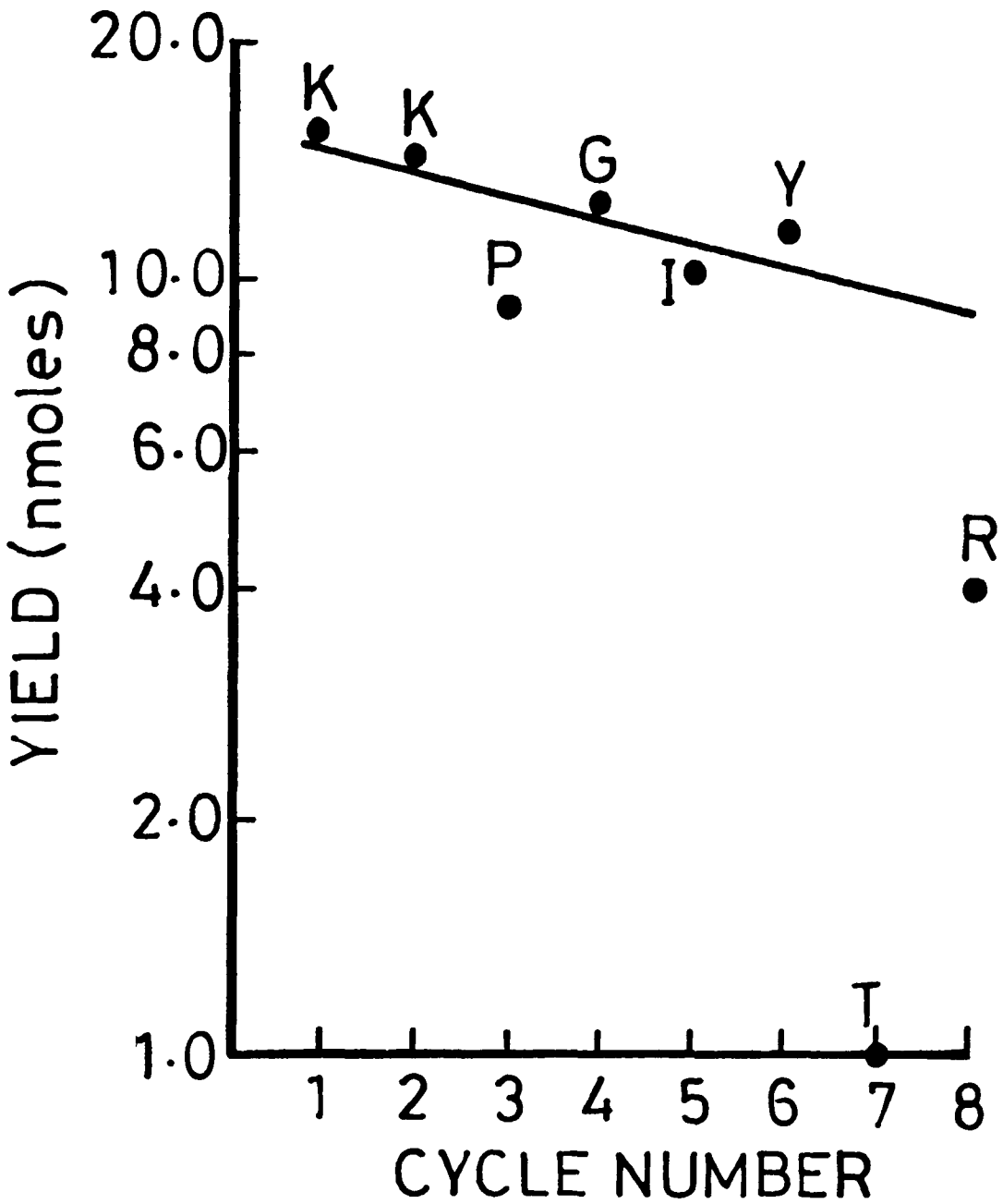


Fig. 4.25 Yield per cycle for automated sequence analysis of peptide T7a-I

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.9. The amount used in the sequencer was 20 nmoles. The stepwise yield calculated from lys 1 and 2 was 95%.

Table 4.12 Amino acid compositions of tryptic peptides from subfractions T7a and T7b

Amino acid	Amino acid composition (mol/mol)				
	T7a-I	T7a-II	T7b-I	T7b-II	T7b-III
Asp	0.9 (1)	1.2 (1)	0.6 (0)	-	0.9 (1)
Thr	1.2 (1)	-	-	-	-
Ser	0.5 (1)	1.2 (1)	-	-	0.6 (1)
Glu	0.9 (1)	0.4 (0)	1.9 (2)	-	0.8 (1)
Pro	1.2 (1)	-	-	-	-
Gly	2.0 (2)	0.9 (1)	-	2.2 (2)	0.9 (1)
Ala	0.5 (1)	3.2 (3)	2.5 (3)	-	0.8 (1)
Val	0.4 (0)	2.0 (2)	-	-	-
Cys	0.4 (0)	-	-	-	0.8 (1)
Met	-	-	-	-	0.6 (1)
Ile	1.4 (1)	0.9 (1)	-	1.0 (1)	0.5 (1)
Leu	0.5 (1)	1.0 (1)	-	1.0 (1)	0.9 (1)
Tyr	1.4 (1)	1.4 (1)	-	-	-
Phe	-	-	-	-	-
His	0.5 (0)	-	1.3 (1)	-	-
Lys	2.4 (2)	-	-	-	0.6 (1)
Arg	2.0 (2)	-	1.0 (1)	1.1 (1)	2.0 (2)

The amino acid compositions are given as mol of residue/mol of peptide. The samples were hydrolysed at 110^{o}C for 48 h. No correction was made for serine and threonine destruction, and tryptophan was not determined. Cystine was measured as SCM-cysteine. -, less than 0.3 mol of residue/mol of peptide.

C-terminal peptide. A sequence of 10 amino acids was obtained (Table 4.11, Fig. 4.26), leaving two amino acids (gly and leu) unaccounted for in the composition. Carboxypeptidase-Y studies were carried out (section 4.6) to establish whether either of these two amino acids were indeed C-terminal in factor $\bar{D}$. Because leu is hydrophobic, it is possible that in a C-terminal position it could be washed from the cup during extractions. By comparing the peptide with other serine proteases, residues 1, 7 and 8 could be aligned with the conserved residues val_{231}, trp_{237} and ile_{238} (CTG numbering).

(3) Peptide T7b-I (yield 19 nmoles) had an amino acid composition (Table 4.12) which was identical to residues 6-12 of intact factor $\bar{D}$. The N-terminal amino acid, determined by dansylation, was glu. Sequence analysis was not carried out.

(4) Peptide T7b-II (yield 10 nmoles) was identical in composition (Table 4.12) to residues 1-5 of intact factor $\bar{D}$ and had N-terminal ile, detected by dansylation. Sequence analysis was not done on this peptide.

(5) Peptide T7b-III (yield approx. 10 nmoles) was identical in amino acid sequence to the secondary peptide which eluted with peptide T7a-I. It contained 9 amino acids and had N-terminal leu (Table 4.11, Fig. 4.27). The sequence from residue 3 onwards was identical to that of CNBr peptide CNBr-2, residues 1-7. Peptide T7a-I was also present at approximately half the concentration of peptide T7b-III.

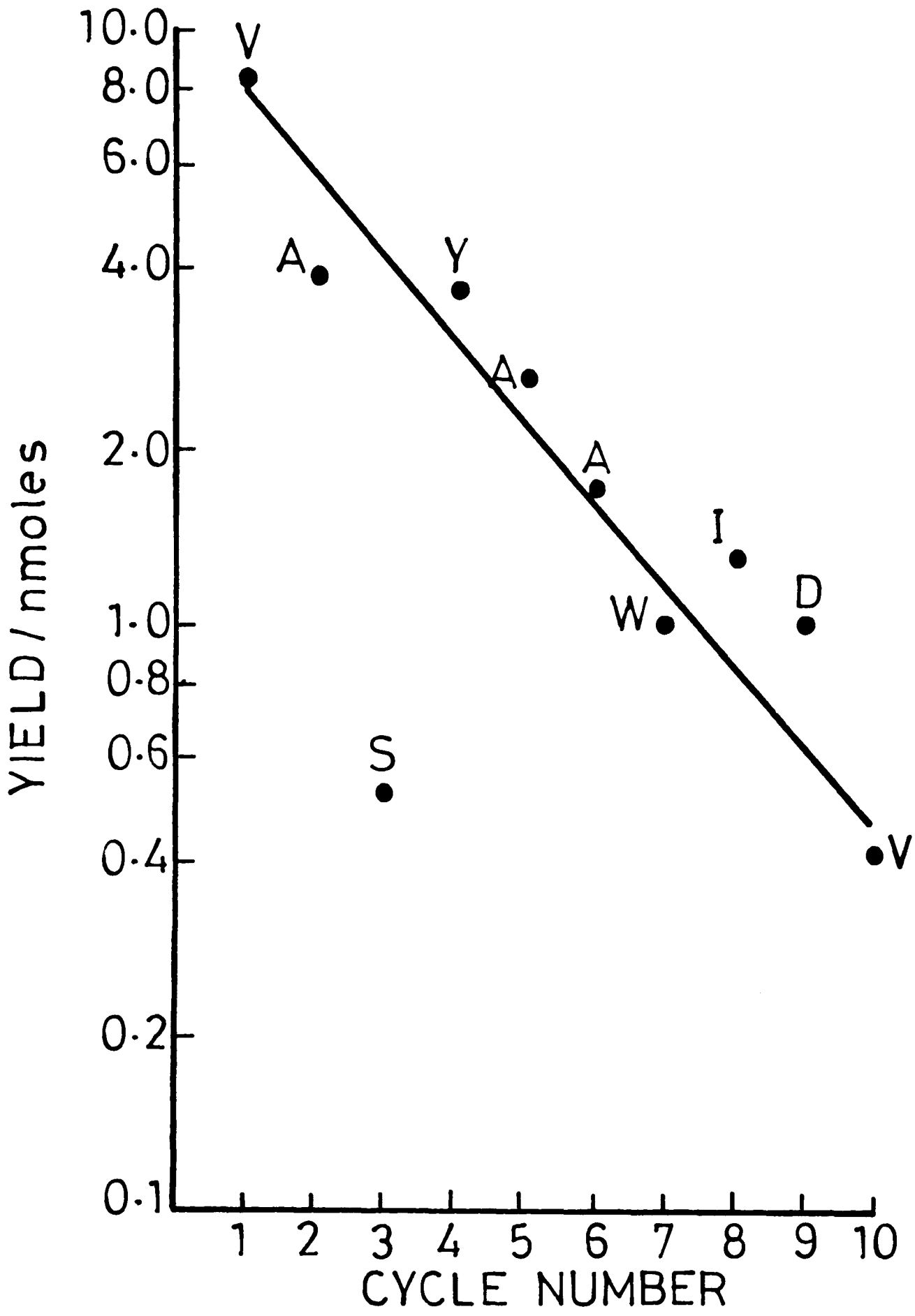


Fig. 4.26 Yield per cycle for automated sequence analysis of peptide T7a-II

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.11. The amount used in the sequencer was approximately 10-15 nmoles. The stepwise yield calculated from ala 2, 5 and 6 was 86%.

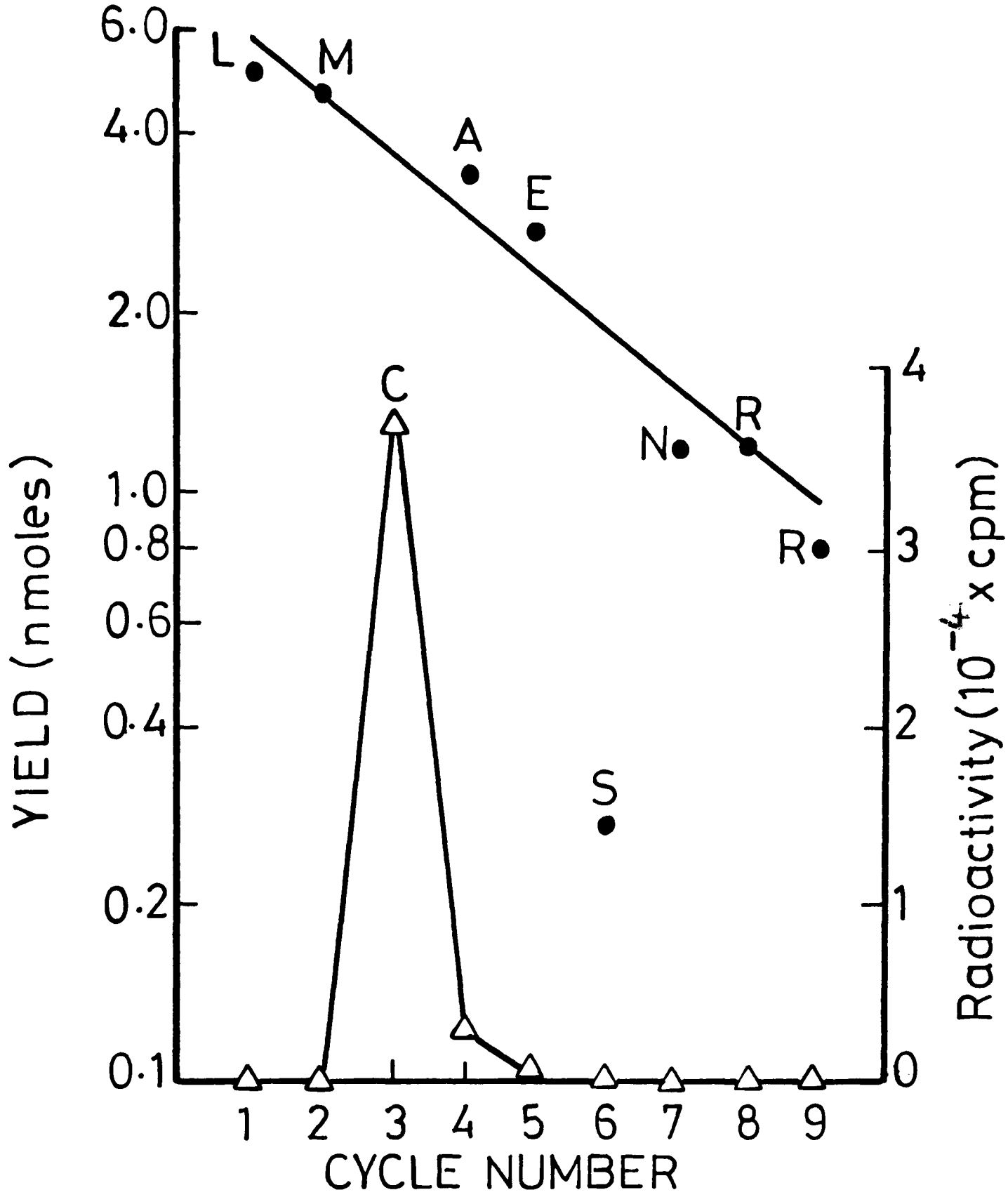


Fig. 4.27 Yield per cycle for automated sequence analysis of peptide T7b-III

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.11. Radioactivity released (Δ) is from SCM-cysteine. The amount used in the sequencer was approximately 10 nmoles. The stepwise yield calculated from arg 8 and 9 was 82%.

The amino acid composition (Table 4.12) reflects the composition of both peptides.

4.5.3.5 Fractions T8 and T9 (mol. wt. <1,000)

(1) One pure peptide was isolated from fraction T8 in a yield of 25 nmoles (Table 13a, Fig. 4.28). This peptide had six amino acids with leu as N-terminal. The composition calculated from the sequence and the amino acid analysis differed only in the possible occurrence of a glycine residue in the amino acid analysis (Table 4.13b). Although the yield was relatively low this could have been caused by aggregation as the peptide contains hydrophobic residues.

The sequence did not show homology with any previously obtained data, but was tentatively aligned by homology with other serine proteases in the region of residues 169-178 (CTG numbering). Residue 169 is often arg or lys, whereas residue 171 (tyr) and 176 (val) are conserved.

(2) Fraction T9 was examined by high voltage electrophoresis at pH 1.9 and was found to have small quantities of ninhydrin-staining material containing arginine (approx. 5 nmoles). No further investigations were carried out on this fraction.

4.5.4 Discussion

Digestion of succinylated factor $\bar{D}$ with trypsin gave 12 major peptides. The yields and approximate mol. wts. are summarised in Table 4.14. The apparently low yields of some peptides may have been caused by their aggregation or,

Table 4.13

(a) Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC. Yields (in nmoles) are given in brackets.

(b) The amino acid composition is given as mol of residue/mol of peptide. The sample was hydrolysed at 110^{o}C for 48 h. No correction was made for serine and threonine destruction, and SCM cysteine, methionine and tryptophan were not determined.

-, less than 0.3 mol of residue/mol of peptide.

Table 4.13 Amino acid composition and amino acid sequence of peptide T8

(a) Amino acid sequence

Residue number	Amino acid found
1	Leu (13.7)
2	Tyr (9.7)
3	Asp (13.6)
4	Val (6.0)
5	Leu (5.3)
6	Arg (2.3)

(b) Amino acid composition

Amino acid	Composition (mol/mol)
Asp	1.2 (1)
Thr	-
Ser	-
Glu	-
Pro	-
Gly	0.6 (1)
Ala	-
Val	1.1 (1)
Ile	-
Leu	1.6 (2)
Tyr	0.8 (1)
Phe	-
His	-
Lys	-
Arg	0.9 (1)

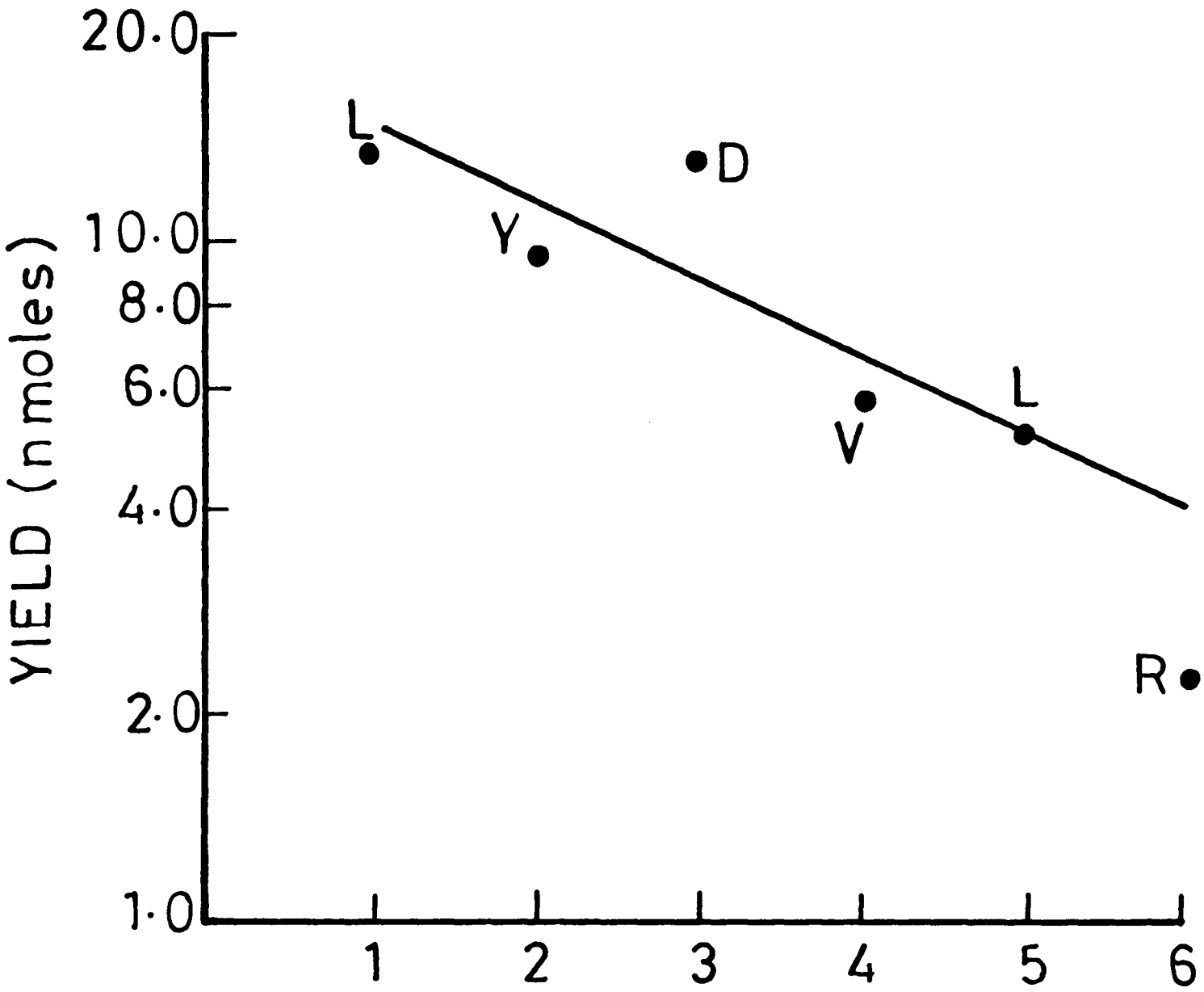


Fig. 4.28 Yield per cycle for automated sequence analysis of peptide T8

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.13a. The amount used in the sequencer was 25 nmoles. The stepwise yield calculated from leu 1 and 5 was 83%.

Table 4.14 Major tryptic peptides of factor $\bar{D}$

Peptide	Approx. mol.wt.	Yield (nmoles)	Method of alignment
T3	5600	15	mol.wt., composition, dansyl
T5a	4300	60	homology
T5b	3900	46	overlap
T6-I	(1100)	52	(homology)
T6-II	2700	46	overlap
T6-III	1700	50	overlap
T7a-I	900	43	homology
T7a-II	1300	10-15	homology
T7b-I	800	19	composition, dansyl
T7b-II	600	10	composition, dansyl
T7b-III	1000	13	overlap
T8	700	25	homology

Molecular weights were estimated either from sequence data or where sequences were incomplete, extrapolation from other data on either factor $\bar{D}$ or other serine proteases. Alignment by homology indicates comparison with other serine proteases, and alignment by overlap indicates comparison with previous data on factor $\bar{D}$. The homology of peptide T6-I with other serine proteases is only tenuous.

in the case of peptides T7b-I and T7b-II, due to incomplete cleavage by trypsin.

Only 0.5% of the protein remained in the 23,000 mol. wt. region on Sephadex G-75 gel filtration. Background hydrolysis in each mol. wt. range, as estimated from minor HPLC fractions, was 10% or less of the yields of major peptides, suggesting that cleavage was >90% complete.

The proposed alignment of tryptic peptides is given in Fig. 4.29 (for the purposes of this discussion, peptide T6-II was assumed to include the tripeptide, val-asp-arg). The method of alignment used for each peptide is given in Table 4.14. The sum of the estimated mol. wts. was 24,600. This is within 5% of the value estimated by polyacrylamide gel electrophoresis of intact factor $\bar{D}$ (section 3.4). From the number of arginine residues in factor $\bar{D}$ (15.0 residues/molecule), sixteen peptides were expected. Although in the alignment given 16 arginine residues are accounted for, in three areas of the molecule the sequence was either uncertain (peptide T6-I) or not obtained (the region between peptides T6-III and T8; and the C-terminal region of peptide T5b). The partial sequence of the missing peptide between peptides T6-III and T8 was obtained from IBA peptides (section 4.4.3.1). It is probably composed of 4 amino acids with the sequence ala-lys-cys-arg. The peptide was probably contained in one of two subfractions of T8 which were not pure following HPLC and have therefore not yet been sequenced. Data is not available as to whether peptide T5b extends as far as peptide T7a-I or whether a further tryptic peptide intervenes. Two

Fig. 4.29

The approximate molecular weights of the peptides are given in Table 4.14. The peptides (with estimated number of amino acids in brackets) are T7b-I (5), T7b-II (7), T3 (approx. 50), T6-I (approx. 12), T5a (approx. 38), T6-II (24), T6-III (15), T8 (6), T7b-III (9), T5b (approx. 35), T7a-I (8) and T7a-II (approx. 12). —, ( ), or a blank space denote positive, tentative or no identification respectively; [ ], peptide only tentatively aligned; |, possible cleavage sites by trypsin.

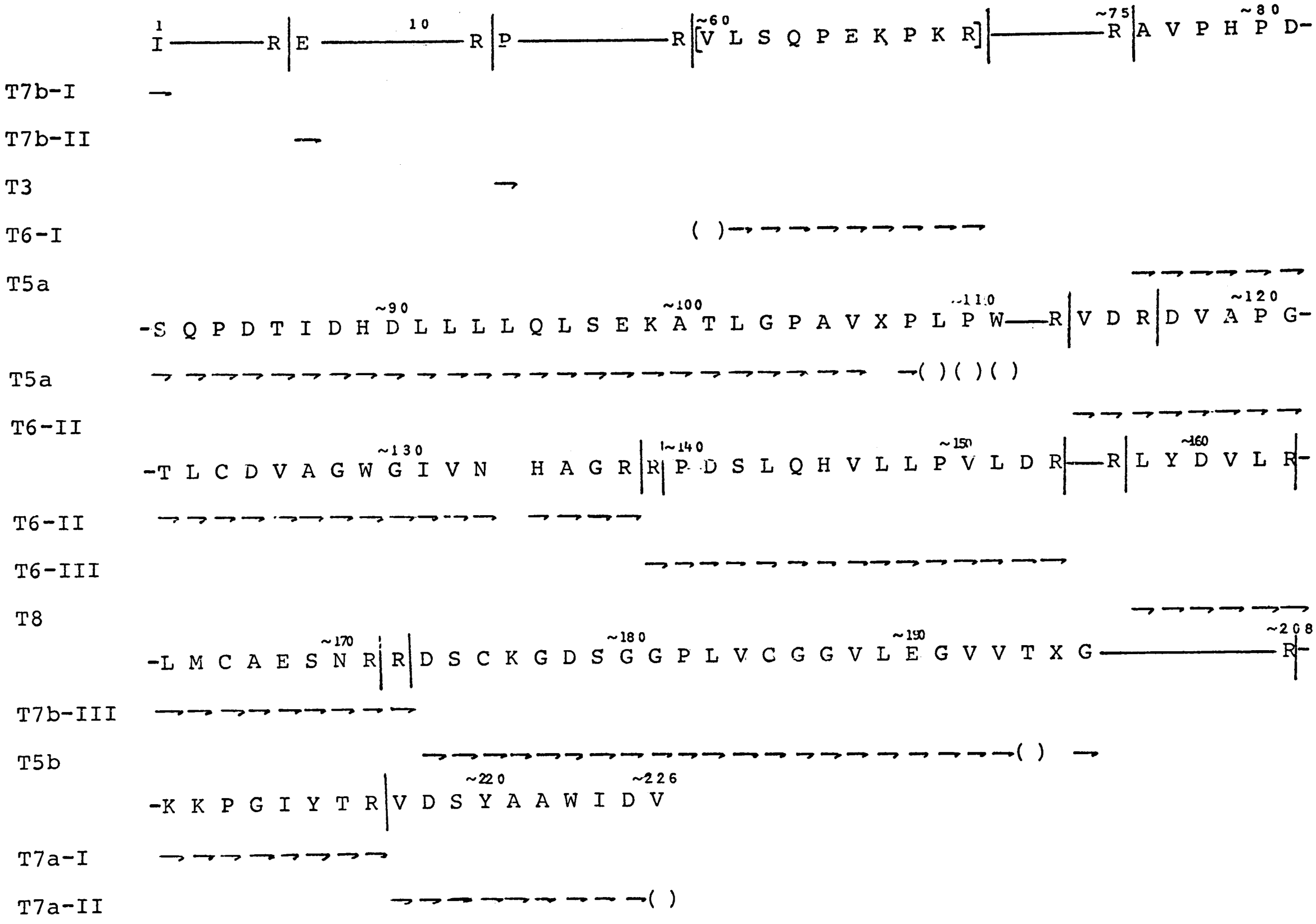


Fig. 4.29 Alignment of major tryptic peptides of factor D

arg-arg sequences were found in factor $\bar{D}$, but significant cleavage at both residues did not occur.

4.6 FURTHER SEQUENCE STUDIES ON FACTOR $\bar{D}$

4.6.1 Introduction

By sequence analysis of the N-terminal of intact factor $\bar{D}$ and of peptides produced by cleavage with cyanogen bromide, o-iodosobenzoic acid and trypsin, 85% or more of the putative amino acid sequence of factor $\bar{D}$ was obtained. However, alignment of many of the peptides, particularly those from trypsin digestion, had been accomplished only by their homology with other serine proteases. This section describes further procedures which were carried out in an attempt to produce overlapping sequences which would confirm the tentative sequence obtained from earlier work. As the sequencing of these peptides has not been completed, the section deals with work still in progress.

The basic approach used in producing overlapping peptides was firstly, to cleave factor $\bar{D}$ with cyanogen bromide, and then to cleave the two larger CNBr peptides further to give overlapping sets of peptides. In the results reported here, peptide CNBr-1 was cleaved using protease from S. aureus V8. V8 protease cleaves specifically after glutamic acid residues when the reaction is carried out in ammonium bicarbonate (Houmard and Drapeau, 1972). However, cleavage is reduced if glu is followed by a bulky hydrophobic residue (e.g. phe, val, leu). Intact factor $\bar{D}$ and peptide CNBr-2 were also cleaved using carboxypeptidase-Y, in order to obtain the C-terminal amino acids. Carboxypeptidase-Y

sequentially removes most amino acids from the C-terminal of protein substrates at pH 5.5-6.5, and is most effective if the penultimate and/or terminal residues are aromatic or aliphatic.

4.6.2 V8 peptides

Peptide CNBr-1 (50-80 nmoles) was isolated as described in section 2.2.11.1, digested with V8 protease and then purified directly by HPLC (section 2.2.11.4). Peptide CNBr-1 contains approximately 13.7 residues of glx, estimated from its size (approximately 140 amino acids) and composition. From previous data, probably about 6 of these residues are glu, so that possibly 7 peptides could be expected. The elution profile obtained by HPLC is given in Fig. 4 30. As glu and gln are not distinguished by amino acid analysis and as gel filtration was not done, estimations of yields and chain length of the peptides by these means were not possible.

Although three of these fractions have been sequenced, one (fraction 4) gave a blank run and a second (fraction 16) gave an ambiguous sequence, suggesting that further purification was necessary. The third (fraction 8) gave a sequence (Table 4.15, Fig. 4.31) which overlapped with and confirmed the N-terminal of tryptic peptide T5a. However, no sequence data was obtained beyond cycle 16 which in peptide T5a was asp. An artefact in the vicinity of the glu peak on the HPLC trace made it impossible to discern whether glu was present in cycle 17. However, the identification of his in this position in peptide T5a was

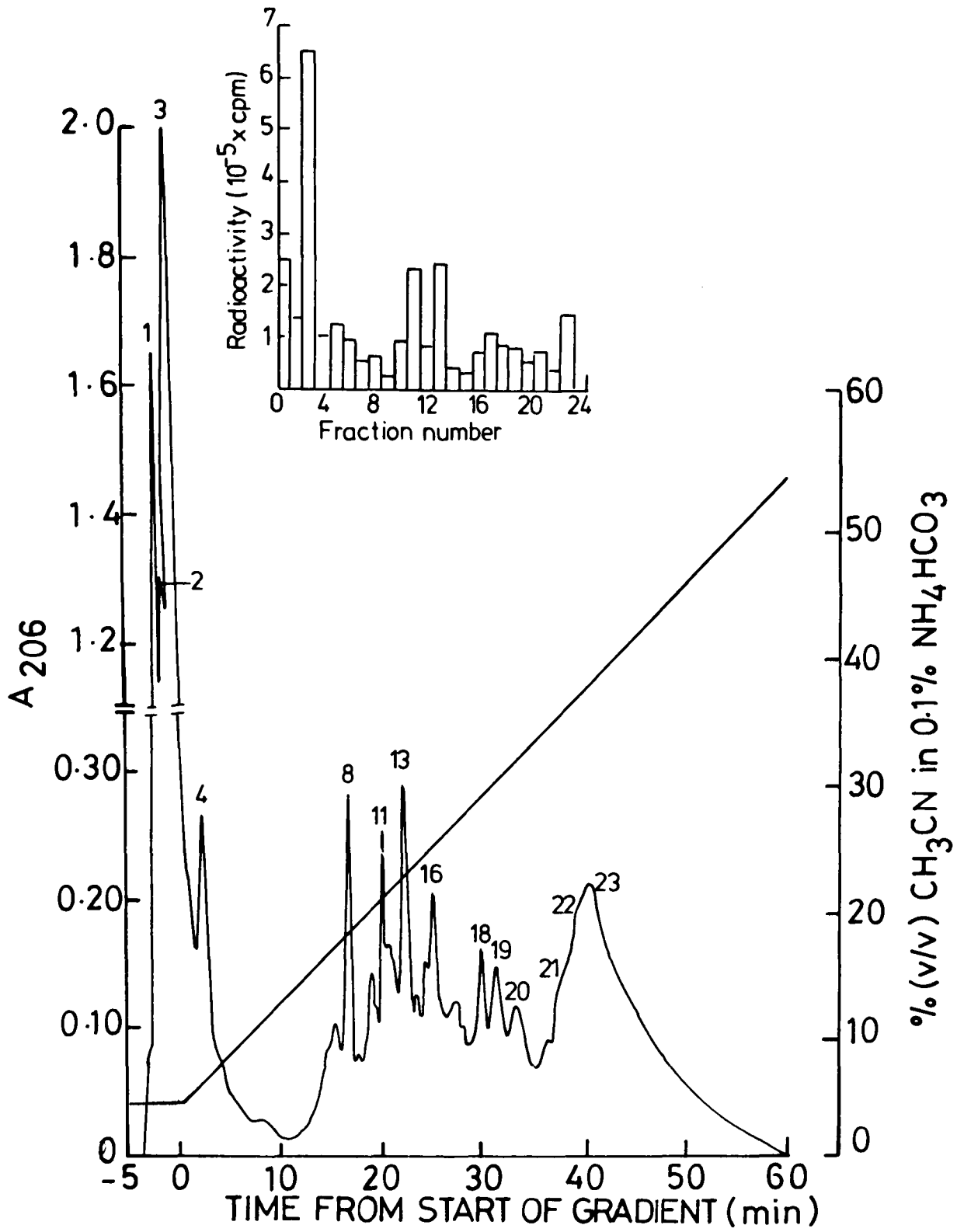


Fig. 4.30 Purification of V8 digest of peptide CNBr-1 by HPLC

The peptides were eluted in a gradient of 5-55% (v/v) CH_3CN in 0.1% (w/v) ammonium bicarbonate. Details are given in section 2.2.11.4.

Inset: samples (1% of total volume) were taken for scintillation counting of ^{14}C-SCM-cysteine.

Table 4.15 Automated sequence analysis of V8 fraction 8

Residue number	Amino acid identified	Yield (nmoles)
1	Val	4.0
2	Leu	3.5
3	Arg	1.7
4	Ala	3.5
5	Val	3.4
6	Pro	1.8
7	His	0.8
8	Pro	1.4
9	Asp	1.8
10	Ser	0.3
11	Gln	1.5
12	Pro	0.7
13	Asp	1.0
14	Thr	0.4
15	Ile	0.4
16	Asp	0.5

Details of the automated Edman degradation are given in section 2.2.12.2. Amino acid PTH derivatives were identified by HPLC.

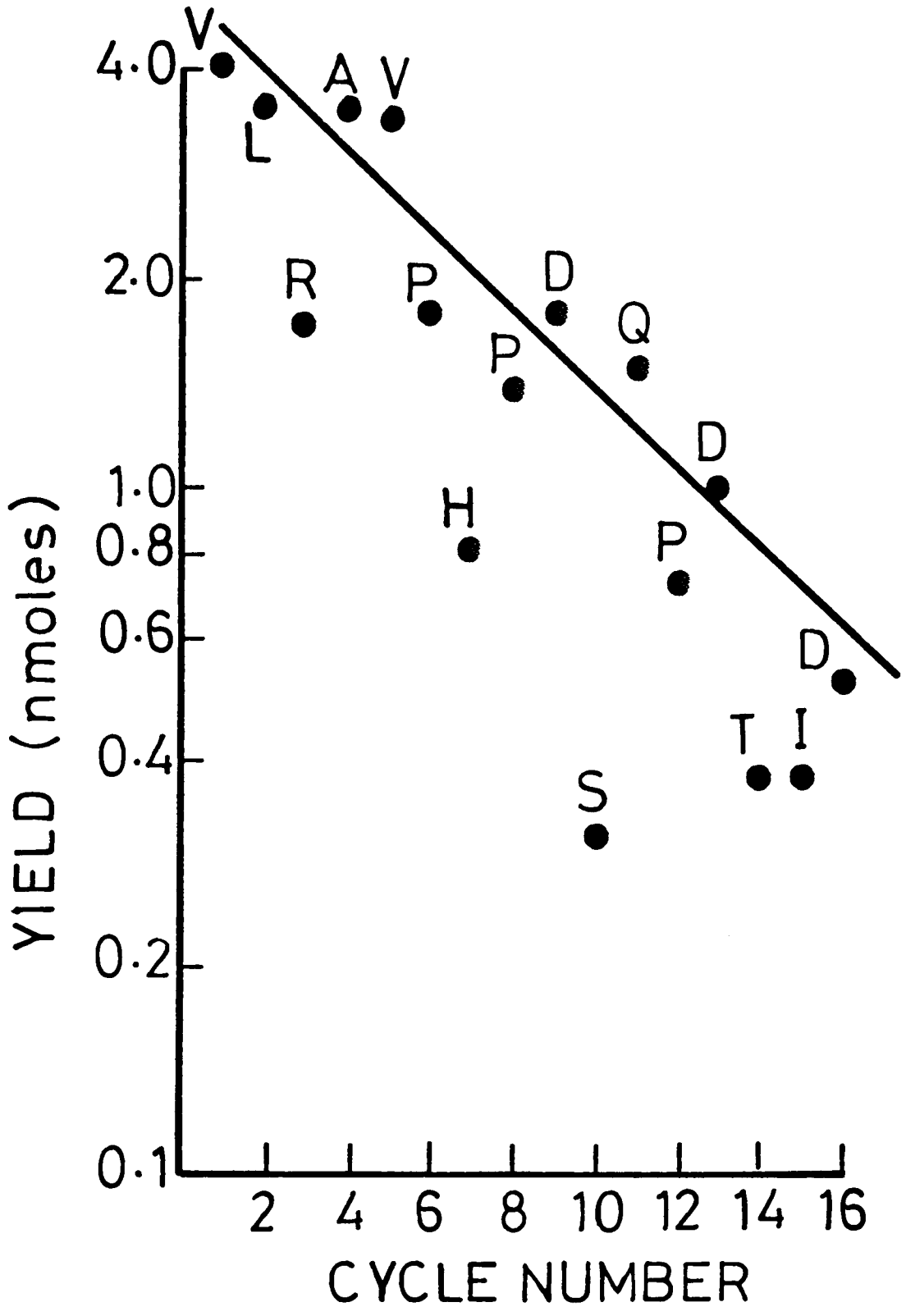


Fig. 4.31 Yield per cycle for automated sequence analysis of peptide V8-8

Amino acids are denoted by the one letter code. The sequence is also found in Table 4.16. The estimated amount used in the sequencer was 13 nmoles. The stepwise yield calculated from pro 6,8 and 12 was 90%.

unequivocal and it is possible that an unexpected cleavage at asp (residue 16) occurred. The composition of the peptide (Table 4.16) correlated well with the sequence data, the only discrepancies being the presence of one glycine residue and the lack of one proline residue in the amino acid analysis. An approximate yield of 13 nmoles was calculated from the composition.

Further purification of other V8 peptides will be necessary before sequencer studies are carried out.

4.6.3 Carboxypeptidase-Y digestion

Carboxypeptidase-Y (CP-Y) digestion was carried out as described in section 2.2.11.5. Experiments were done in which intact factor $\bar{D}$ was digested for 5 h, or peptide CNBr-2 was digested for time periods of 0-1 h, 5 h and 16 h. Data from digests incubated for longer than 1 h were not easily interpreted. However, after 1 h, the amino acids released from peptide CNBr-2 (Table 4.17) suggested a C-terminal peptide which was similar in composition to tryptic peptide T7a-II, residues 3-10. In addition leu, which was present in the composition but not in the sequencer run of peptide T7a-II was also released. The relative rates of release of amino acids (Fig. 4.32) showed that leu was the most rapidly released amino acid followed by ala, val, asp, ile and tyr in that order (ser and trp were not evaluated). These rates of release correlated with those expected from the sequence of peptide T7a-II, except that alanine was released much earlier than expected. The most likely explanation of this discrepancy is the occurrence

Table 4.16 Amino acid composition of V8 fraction 8

Amino acid	Composition (mol/mol)
Asp	3.0 (3)
Thr	0.7 (1)
Ser	0.7 (1)
Glu	1.5 (1)
Pro	1.8 (2)
Gly	1.2 (1)
Ala	1.3 (1)
Val	1.8 (2)
Cys	-
Met	-
Ile	0.6 (1)
Leu	1.2 (1)
Tyr	-
Phe	-
His	1.2 (1)
Lys	0.5 (0)
Arg	0.8 (1)

The amino acid composition is given as mol of residue/mol of peptide. The sample was hydrolysed at 110^{o}C for 48 h. No corrections were made for serine and threonine destruction, and tryptophan was not determined. Cystine was measured as SCM-cysteine. -, less than 0.3 mol/mol.

Table 4.17 Amino acids released from peptide CNBr-2 by digestion with CP-Y for 1 h.

Amino acid	Nmoles released
Asp	0.5
Thr	-
Ser†	1.2
Glu	-
Pro	-
Gly	-
Ala	2.0
Val	1.0
Ile	0.4
Leu	1.4
Tyr	0.8
Phe	-
His	-
Lys	-
Arg	-

Following digestion with CP-Y for 1 h, samples (3-5 nmoles) were boiled for 5 min at 100°C, freeze-dried and analysed directly. No corrections were made for destruction of serine and threonine. Cystine, methionine and tryptophan were not estimated. Amounts released were corrected for background (measured at 0 h). -, less than 0.3 nmoles released.

† gln and asn were not distinguishable from ser in this analysis.

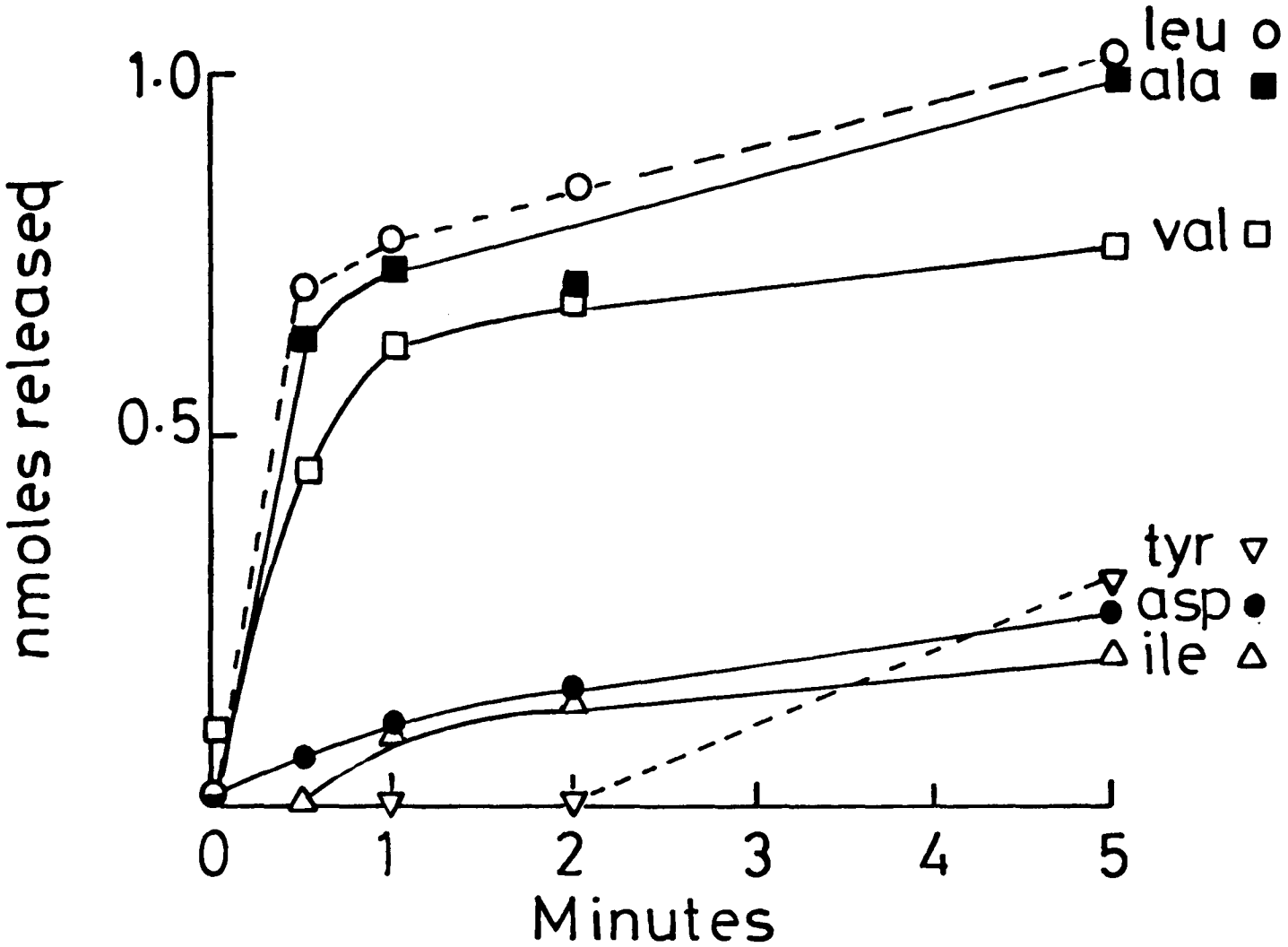


Fig. 4.32 Release of amino acids from peptide CNBr-2 by digestion with CP-Y

Peptide CNBr-2 (approx. 3-5 nmoles) was incubated with CP-Y at 25°C over time periods from 0-5 min. Samples were boiled for 5 min to inactivate CP-Y, then analysed after freeze-drying. Release of tryptophan and serine were not estimated.

of three alanine residues close to the C-terminal.

From these results and from the sequence and composition of peptide T7a-II, it was concluded that the C-terminal amino acid residue of factor $\bar{D}$ was leucine.

4.7 CONCLUDING SUMMARY

In the studies reported here, approximately 88% of the amino acid sequence was established with approximately 84% of the identities being unambiguous (Fig. 4.33). The alignment of peptides was deduced largely by homology with other serine proteases, but where possible by overlaps between different sets of peptides. Further studies to confirm the alignment and to obtain the two "missing" regions of sequence (from residues 72-84 and 212-221, CTG numbering) will be necessary. Unfortunately, one of the missing sequences spans the putative secondary substrate binding site (CTG residues 214-216).

By comparison of the approximate composition of factor D from sequence data and from amino acid analysis (Table 3.3), and allowing for gaps in the sequence, the most obvious discrepancies were in the estimated numbers of some hydrophobic residues: 22 valines in the sequence compared with approximately 17-18 in the amino acid analysis, and 27 leucines compared with 22-23 in the amino acid analysis. It is probable that this discrepancy arose from incomplete release of these amino acids by acid hydrolysis. Due to the limited amounts of factor $\bar{D}$ available, hydrolyses were normally carried out for 48 h only whereas complete release of leucine and valine requires hydrolysis for 72-96 h. The estimated number of tryptophans for the sequence data was 4, compared with 3 from amino acid analysis. No phe has yet been detected in the sequence, although the composition suggests that 3 phe residues may occur. It is

Fig. 4.33 Partial amino acid sequence of factor $\bar{D}$

The sequence of factor $\bar{D}$ is given together with the peptides from which the sequence was deduced. Peptides have been aligned with other serine proteases and residues are numbered according to chymotrypsinogen numbering. Gaps (-) have been left to allow maximum homology with other serine proteases. Starred residues (*) are invariant in other serine proteases. The sequence is incomplete from residues 72-84 and 212-221. Peptide T6-I may be derived from the segment between residues 72 and 84 but no evidence is yet available.

```
16                20                30                40                50
I L G G* R E A E A H A R P* Y M A S V Q L - - ·· - N(G)A E L C* G* G V L V A E Q W V L T A*
└──────CNBr-3──────┘└──────────────────────────────────────────────────────CNBr-1──
└──────────────────────────IBA-I──────────────────────────┘└IBA-II─
└T7b-II┘└──T7b-I──┘└──────────────────────────────────────────────T3──────

*  *    60                                  70        85        90
A X C L E D A A G G(Y)V X - - - - - - - V L L G A I(──────)E V L R A V P H P D S Q -
──────────────────────────────CNBr-1──────────────────────────────────────────────
──────────────────────────────IBA-II──────────────────────────────────────────────
──────────T3──────────┘                                  └──T5a──────
                                                  └──V8-8──────

      100               110               120               130
P D T I D - - - H D* L L L L Q L* S E K A T L G P A V X P - -(L* P W)G R V D R D V A P G
──────────────────────────────CNBr-1──────────────────────────────────────────────
──────────IBA-II──────────┘                                  └──IBA-III──
──────────T5a──────────────────────────────────────┘└──T6-II──
──V8-8──┘

      140*  *           150               160*  *           170
T L C D V A G W G I - - - - V N H - A G R R - P D S L* Q H V L L P* V L D R A K C* R - L
──────────────────────────CNBr-1──────────────────────────────────────────────
──IBA-III──┘└──────────────────────IBA-IV──────────────────────────────
──────T6-II──────┘└──────T6-III──────┘  └T8-

      180               190*  * * * * * * 200
Y D - - - V L R L M - C A E S - - - - N R R - D S C K C D S G G P L V - - C G G V L E
──CNBR-1──┘└──────────CNBR-2──────────────────────────
──────────────────IBA-IV──────────────────────────────
──T8──┘└──T7b-III──┘└──────T5b──────

      210*  222               230               240
G V V(T)X G(──────────────)R K - K P G I Y T R V A S Y A A W I D(V)(L)
──────────CNBr-2──────────┘
──────IBA-IV──────────────────────────┘└IBA-V──┘
──────T5b──────┘└──T7a-I──┘└──T7a-II──┘
```

possible that one of these will be found in the secondary substrate binding site (CTG residue 215) where trp usually occurs, although in elastase this is replaced by phe, and in rat group specific protease (Woodbury et al, 1978), by tyr. The probability that factor $\bar{D}$ does not have trp in this site is also suggested by the lack of an IBA peptide which showed homology with this region of serine proteases. At least one phe is also common in serine proteases in the region of CTG residues 72-84, the other missing sequence in the data reported here. Of the eight half-cystine residues present in factor $\bar{D}$, seven were detected, and by analogy with other serine proteases the eighth will probably be found between residues 212-221 (CTG numbering).

A comparison of the sequence data with other serine proteases is found in Chapter 6.

CHAPTER 5

ENZYMIC ACTIVITY OF FACTOR $\bar{D}$

5.1 INTRODUCTION

Factor $\bar{D}$ was recognised as a serine esterase when Fearon et al (1974) discovered that pretreatment of factor $\bar{D}$ with iPr_2P-F completely inhibited the ability of factor $\bar{D}$ to generate the haemolytically active cellular intermediate, $EAC\overline{43B}$. In the same study, Fearon et al described a precursor form of factor $\bar{D}$ which was resistant to inhibition by iPr_2P-F and could be activated by trypsin. Lesavre and Müller-Eberhard (1978), using the same fractionation procedure for obtaining "precursor factor $\bar{D}$" as Fearon et al, measured factor $\bar{D}$ by both its haemolytic activity and by radioimmunoassay of ^{125}I-labelled factor $\bar{D}$ but could find no evidence for a precursor form of the enzyme.

Reports on the ability of factor $\bar{D}$ to utilise synthetic substrates have also been conflicting. Earlier reports had suggested that factor $\bar{D}$ could utilise certain synthetic lysyl and arginyl esters (Dieminger et al, 1976; Volanakis et al, 1977). More recent reports have not confirmed these early findings (Lesavre and Müller-Eberhard, 1978; Davis et al, 1979b).

The main aims of the studies presented here were (1) to repeat the experiments of Fearon et al (1974) to see whether fresh serum contained any trypsin inducible factor $\bar{D}$ haemolytic activity; (2) to review the activity of factor $\bar{D}$ towards synthetic ester substrates and to extend the studies to include the thioester , thiobenzyl-benzoxycarbonyl-lysinate

(ZLSB), synthetic nitroanilide substrates and the active site titrant, p-nitrophenyl-p'-guanidinobenzoate (NPGB);
(3) to test the sensitivity of factor $\bar{D}$ to a variety of inhibitors of both serine and non-serine proteases, as well as against some common plasma inhibitors.

5.2 ASSAY FOR FACTOR $\bar{D}$ ACTIVITY

Haemolytic assays for factor $\bar{D}$ were carried out as described by Martin et al (1976), using 10 µl samples. In the studies reported in this chapter, all assays were performed in duplicate, and standardised as described in section 2.2.3. Fig. 5.1(a) shows a typical standard plate and Fig. 5.1(b) a typical plot of area v. standard serum concentration. The haemolytic activity of factor $\bar{D}$ in the assay increases roughly twofold with increasing pH of sample buffer from pH 5.0 to pH 8.0. Buffer composition does not seem to affect haemolytic activity except in the cases of (1) high concentrations of phosphate salts ($\geq$ 0.25 M) in which the apparent activity of factor $\bar{D}$ may be reduced, supposedly due to local sequestration of magnesium ions, and (2) nucleophilic buffers such as Tris which enhance the enzyme activity presumably by increasing the rate of decay of the acyl enzyme (section 1.3.2.2.).

Fig. 5.1 Assay for factor $\bar{D}$ haemolytic activity

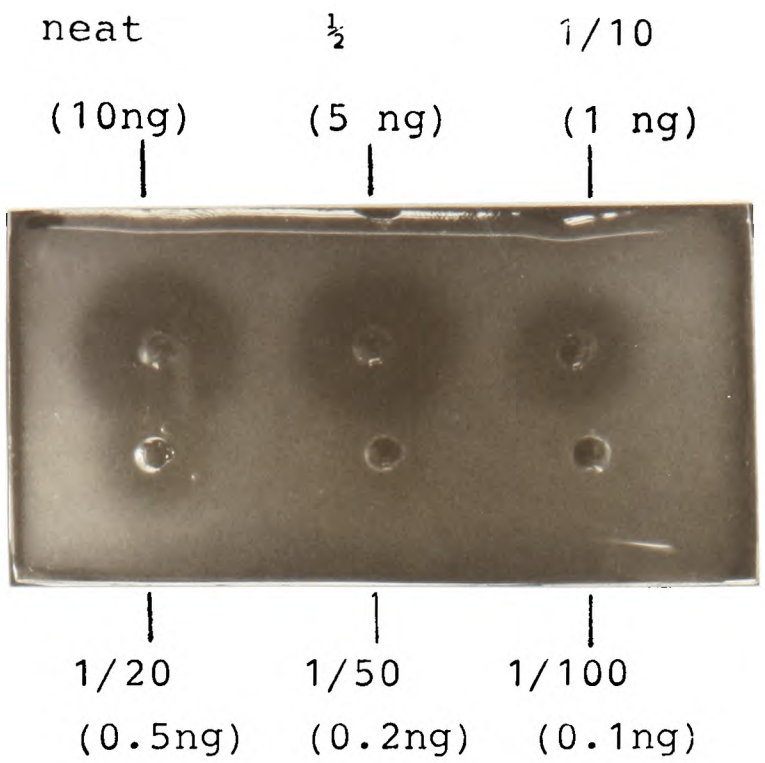
(a) Factor $\bar{D}$ haemolytic diffusion plate

Dilutions of a pooled normal human serum (NHS) (10 µl) were added to wells (3 mm diameter) in the assay plate and allowed to diffuse at 4°C overnight. Haemolysis was developed at 37°C. Details are given in section 2.2.3. The dilutions and approximate factor $\bar{D}$ concentrations are indicated by each well.

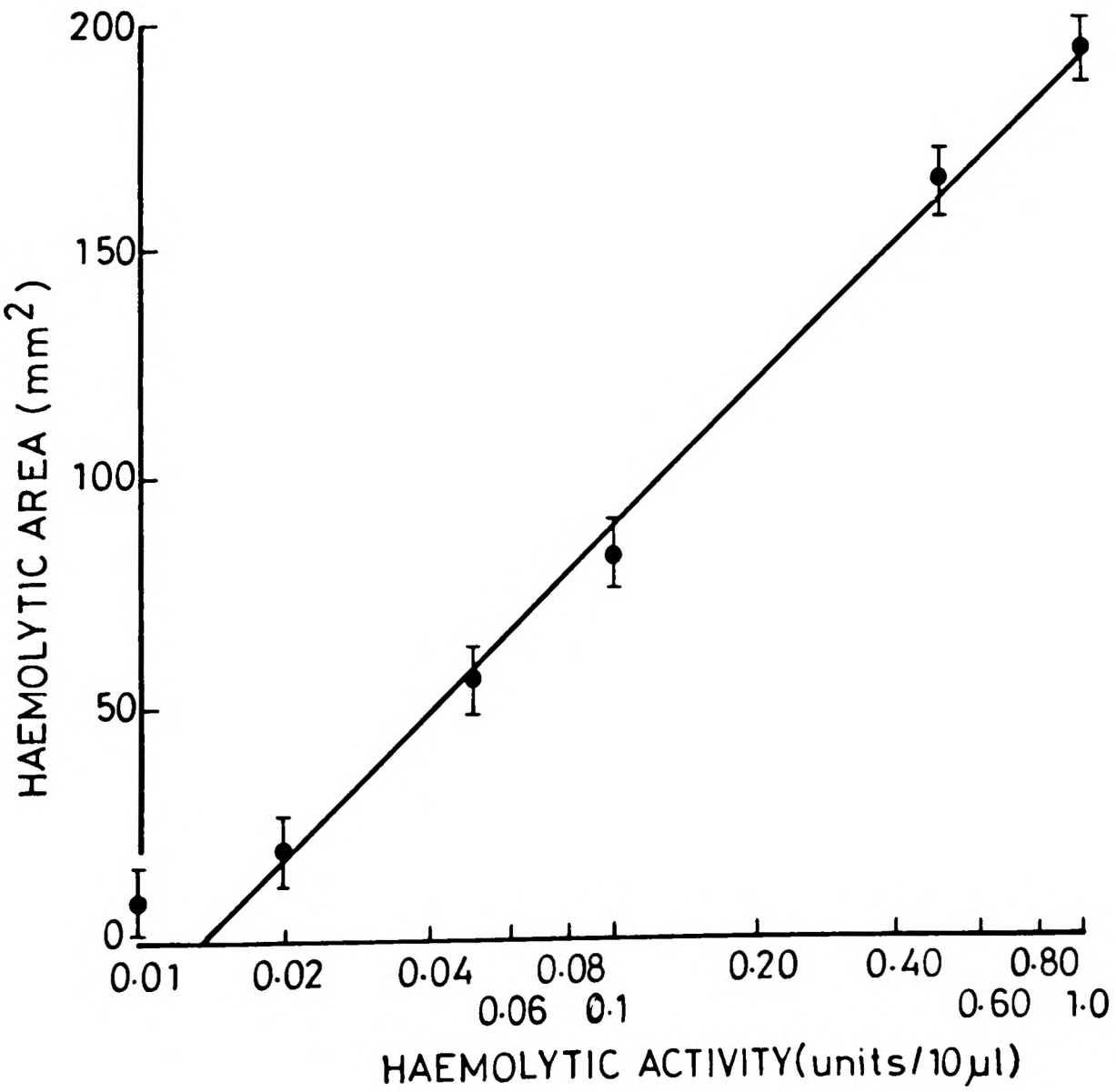
(b) Standard plot for haemolytic diffusion plate assay

Haemolytic areas were calculated from assay plates (Fig. 5.1a). One unit of activity is defined as the haemolytic area obtained using 10 µl of undiluted pooled NHS. The standard deviation at each point is indicated by bar-lines.

(a)


neat
½
1/10
(10ng)
(5 ng)
(1 ng)
1/20
1/50
1/100
(0.5ng)
(0.2ng)
(0.1ng)

(b)


200
150
100
50
0
HAEMOLYTIC AREA (mm²)
0·01
0·02
0·04
0·06
0·08
0·1
0·20
0·40
0·60
0·80
1·0
HAEMOLYTIC ACTIVITY(units/10µl)

5.3 TRYPSIN-INDUCIBLE FACTOR $\bar{D}$ ACTIVITY

5.3.1 Pilot study

Crude factor $\bar{D}$ was isolated from the 25,000 - molecular weight fraction during the preparation of factor $\bar{D}$-deficient serum on Sephadex G-75 (section 2.1.5.2). The crude factor $\bar{D}$ (1 ml containing approximately 2 µg) was incubated with iPr_2P-F (1 mM) for 30 min at 30°C. The inhibited sample was dialysed overnight to remove excess iPr_2P-F and the inactivated factor $\bar{D}$ (0.2 µg) assayed for trypsin inducible haemolytic activity (section 2.2.8.1). The results (Table 5.1) indicated that

(1) 1 mM iPr_2P-F was insufficient to completely inhibit factor $\bar{D}$ haemolytic activity. This was in contrast to the observation of Fearon et al (1974) who observed complete inhibition at this concentration.

(2) addition of up to 10 µg/ml of trypsin had no enhancing effect on the haemolytic activity of iPr_2P-F-inhibited factor $\bar{D}$.

Since the assay plate contains all factors necessary for the activation of the alternative pathway, except for factor $\bar{D}$, it will presumably also contain any putative activator of factor $\bar{D}$. It is conceivable, therefore, that the residual activity obtained when iPr_2P-F-inhibited factor $\bar{D}$ was assayed without trypsin activation, was due to activation of "precursor factor $\bar{D}$" in the assay plate itself. However, factor $\bar{D}$ could be completely inhibited using 5-10 mM iPr_2P-F (section 5.5), so it is unlikely that the residual activity observed in this study was due to a precursor form of factor $\bar{D}$.

Table 5.1 Trypsin-inducible haemolytic activity of factor $\bar{D}$ - pilot study using crude factor $\bar{D}$

Trypsin added (µg/ml)	Haemolytic activity (units/10 µl)	
0	0.25	0.09
1	0.25	0.08
2	0.16	0.09
4	0.16	0.08
8	0.16	0.08
10	0.16	0.05
Control	0	0

Crude factor $\bar{D}$ was inhibited with 1 mM iPr_2P-F, dialysed, then "activated" with trypsin. Details are given in the text. The results are from two separate experiments. The control sample contained SBTI-inhibited trypsin (10 µg/ml) in assay buffer.

Factor $\bar{D}$-deficient serum, R(D), which lacks factor $\bar{D}$ haemolytic activity when freshly prepared, was observed to cause haemolysis in assay plates after storage for several weeks at 4°C or at -70°C. Therefore, freshly prepared R(D) was inhibited with iPr_2P-F and then activated with trypsin as described for crude factor $\bar{D}$. However, no haemolytic activity developed following treatment with trypsin at a concentration of up to 10 µg/ml.

5.3.2 QAE-Sephadex A-50 chromatography of plasma

Plasma was fractionated on a column of QAE-Sephadex A-50 (section 2.2.4.1) and the column fractions assayed for factor $\bar{D}$ haemolytic activity and trypsin-inducible factor $\bar{D}$ activity, using trypsin at a concentration of 5 µg/ml (Fig. 5.2). The peak of factor $\bar{D}$ haemolytic activity eluted at 15 mS. This was at a considerably higher conductivity than was found by Fearon et al, 1974 (8.5 mS), although the overall elution profile was similar. A shoulder of haemolytic activity at 18 mS was shown to be positive also for factor B haemolytic activity. In two separate experiments a slight shoulder of haemolytic activity was found which eluted at approximately 12 mS, in a relative position equivalent to the position of the "precursor factor $\bar{D}$" of Fearon et al (1974). However, trypsin treatment of these fractions without prior inactivation by iPr_2P-F did not enhance the haemolytic activity further. Because of the low concentration of factor $\bar{D}$ in the column fractions, the "precursor factor $\bar{D}$" and factor $\bar{D}$ fractions were separately pooled, concentrated, and trypsin activation repeated before

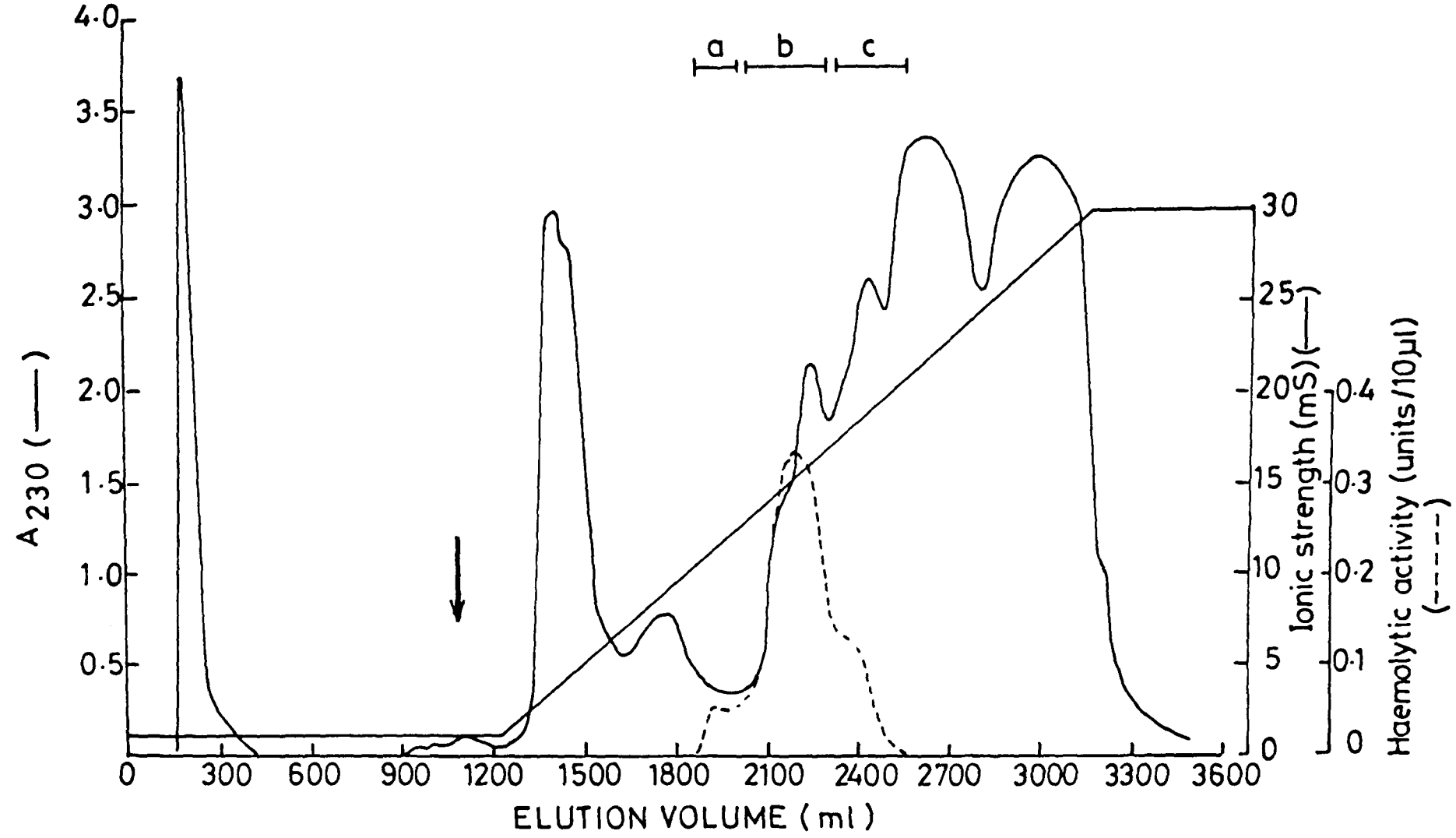


Fig. 5.2 Ion-exchange chromatography of plasma on QAE-Sephadex A-50

Plasma pseudoglobulin in 0.0035 M sodium phosphate, pH 8.0 was fractionated on QAE-Sephadex A-50 as described in section 2.2.4.1. The arrow marks the start of the salt gradient. Fractions marked a, b and c are "precursor factor $\bar{D}$", factor $\bar{D}$ and factor B containing pools respectively.

and after inhibition with 2 mM iPr_2P-F. In addition, the trypsin activation of iPr_2P-F-inhibited samples was carried out with and without the dialysis step to remove excess iPr_2P-F. The results are summarised in Table 5.2. More than 75% of the "precursor factor $\bar{D}$" haemolytic activity was inhibited by iPr_2P-F. Furthermore, no significant enhancement of haemolytic activity was observed following treatment with iPr_2P-F and trypsin. The apparent slight increase in haemolytic activity observed when "precursor factor $\bar{D}$" was treated with trypsin (10 µg/ml) was not regarded as significant, as this degree of variation could also be accounted for by experimental error.

5.3.3 Conclusions

In the experiments described here, no evidence was found for the presence of a precursor form of factor $\bar{D}$ which was insensitive to iPr_2P-F and activable by trypsin. Trypsin did not significantly enhance the haemolytic activity of factor $\bar{D}$ before or after treatment with iPr_2P-F. In fact at a concentration of 10 µg/ml, trypsin seemed to have a slight inhibitory effect on haemolytic activity (Tables 5.1 and 5.2). Interestingly, this inhibitory effect was not observed in the "precursor factor $\bar{D}$" pool. The occurrence of a shoulder of haemolytic activity at 12 mS (equivalent in elution position to the precursor factor $\bar{D}$ of Fearon *et al*) was interesting, however, and it is possible that this enhancement was caused by another plasma protein which affects the alternative pathway. No attempts were made to identify the protein, but in view of the symmetrical distribution of factor $\bar{D}$ antigen found by Lesavre and Müller-Eberhard (1978), it seems unlikely that it is related to factor $\bar{D}$.

Table 5.2 Trypsin-inducible haemolytic activity of factor $\bar{D}$ - fractions from chromatography of fresh frozen plasma on QAE-Sephadex A-50

Fraction	Trypsin added	Haemolytic activity (units/10 µl) of samples treated with		
	µg/ml	No iPr_2P-F	2 mM iPr_2P-F Dialysed	2 mM iPr_2P-F No dialysis
"Precursor Factor $\bar{D}$"	0	0.42	0.12	0
	5	0.42	0.10	0
	10	0.47	0.11	0
Factor $\bar{D}$	0	0.50	0.05	0.02
	5	0.42	0.05	0.02
	10	0.25	0.05	0.05

The active fractions from QAE-Sephadex A-50 were pooled and concentrated by Diaflo-ultrafiltration. The fractions were pretreated where applicable with 2 mM iPr_2P-F and duplicate samples were then "activated" with trypsin. Details are given in the text.

5.4 ENZYME ACTIVITY OF FACTOR $\bar{D}$ WITH SYNTHETIC SUBSTRATES

5.4.1 p-Nitroanilide substrates

In a pilot study, factor $\bar{D}$ obtained from the peak fraction of a cellulose CM-32 column (see purification procedure, section 2.2.4.3) was used to test the amidolytic activity of factor $\bar{D}$ towards 14 p-nitroanilide substrates. The substrates used, together with their code numbers by which they will be referred in the text, are listed in Table 2.1. The assays were carried out as described in section 2.2.6.

Of the 14 substrates tested, only three (S2393, S2407 and S2238) were hydrolysed to any extent. The results for both factor $\bar{D}$ and other serine proteases are summarised in Table 5.3. Interestingly, substrate S2222 which others have found to be slowly hydrolysed by factor $\bar{D}$ (Lesavre and Müller-Eberhard, 1978; Davis et al, 1979b) was not hydrolysed in this study.

Because several proteins including factor $\bar{D}$ elute close together in the gradient of the CM-cellulose 32 column, a preparation of factor $\bar{D}$ was carried out using plasma as the starting material and the fractions eluting from 16-36 mS on CM-cellulose 32 were tested for amidolytic activity towards the substrates S2393, S2407 and S2238 (Fig. 5.3). The study showed that while the glycoprotein eluting at 25-28 mS did not hydrolyse the substrates to any extent, they were rapidly hydrolysed by the protein eluting at 18-22 mS. In addition, a mixture of anti-thrombin III (at a final concentration of 0.2 units/ml) and SBTI (at a final concentration of 0.7 mg/ml) caused a slight decrease in this activity to about 80% of an un-inhibited control. The experimental error in ΔA_{405}

Table 5.3 Amidolytic activity of factor $\bar{D}$ and other serine proteases

Enzyme	p-nitroaniline formed (10^3 x µmol/min/mg) using		
	S2407	S2393	S2238
Trypsin	177,000	92,300	100,000
Chymotrypsin	0.574	0	0
Factor $\bar{D}$	2.96	2.95	3.93
$C\bar{1}r$ †	3.63	3.75	1.55
$C\bar{1}s$ †	33.6	49.6	1.06

† A.W. Dodds (unpublished results)

The release of p-nitroaniline was measured spectrophotometrically at 405 nm (section 2.2.6.1). The concentrations of enzyme used were : trypsin, 1×10^{-9} M; chymotrypsin, 1×10^{-7} M; factor $\bar{D}$, 6.5×10^{-8} M; $C\bar{1}r$, 5×10^{-7} M; $C\bar{1}s$, 5×10^{-7} M. The composition of substrates S2407, S2393 and S2238 are found in Table 2.1.

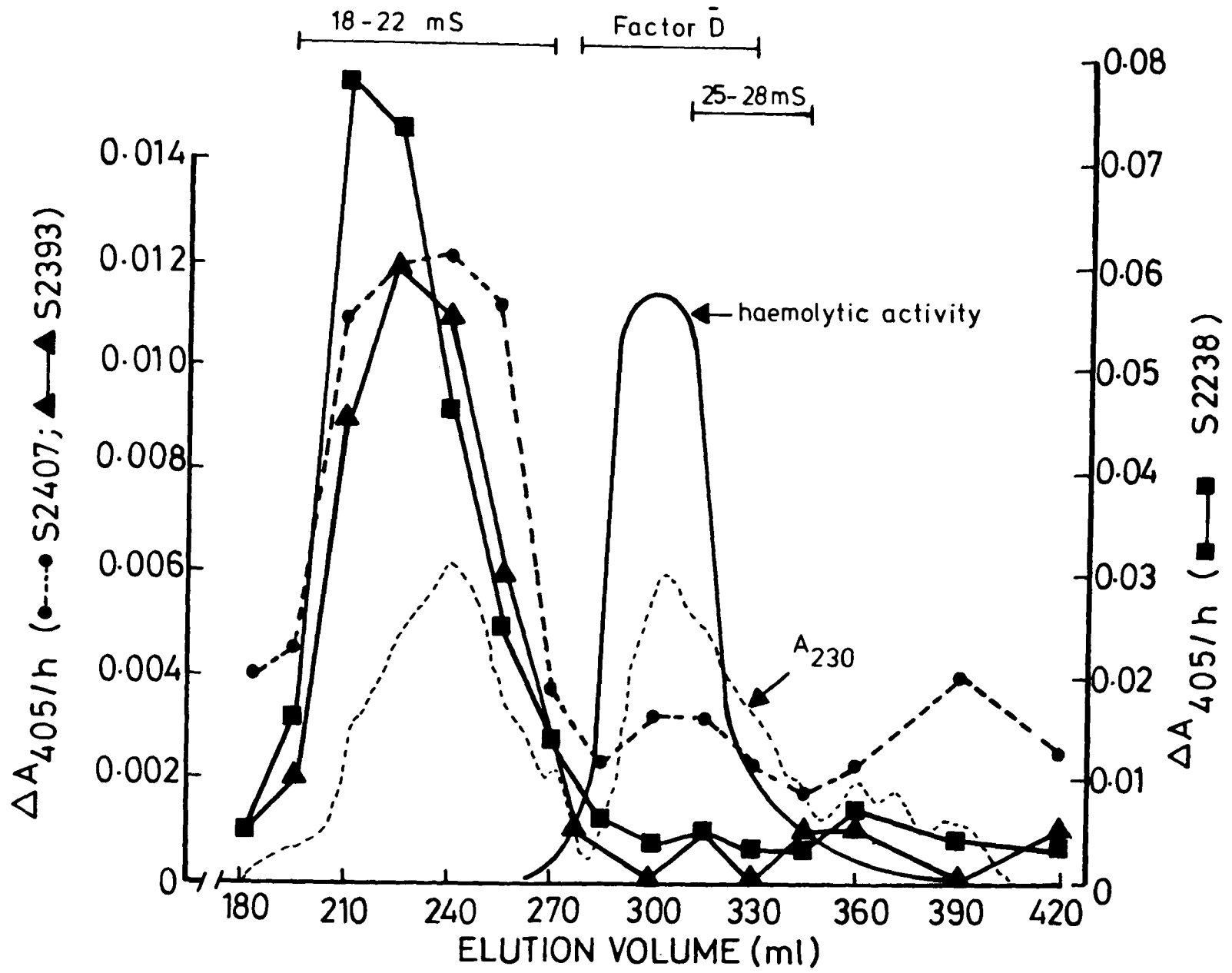


Fig. 5.3 Amidolytic activity of fractions obtained by CM-cellulose 32 chromatography of crude factor D̄

Plasma pseudoglobulin was fractionated on CM-Sephadex C-50, Sephadex G-75 and CM-cellulose 32 as described in section 2.2.4.3. Effluent eluting between 16 and 36 mS on the salt gradient was collected in 3 ml fractions, and samples (100 µl) were assayed spectrophotometrically as described in section 2.2.6.

for substrates S2238 and S2393 was $\pm$ 0.001, and for substrate S2407, $\pm$0.002. Thus, it could not be distinguished whether the small amount of activity observed in factor $\bar{D}$-containing fractions was due to factor $\bar{D}$ itself, traces of the 18-22 mS protein, or experimental error.

5.4.2 Arginyl and lysyl esters

The esterolytic activity of factor $\bar{D}$ obtained from the peak of the haemolytic activity on CM-cellulose 32 (section 2.2.4.3) toward TAME, BAEE and ZLSB was measured as described in section 2.2.6. The results are summarised in Table 5.4.

Factor $\bar{D}$ showed no esterolytic activity towards TAME or BAEE. These results were in contrast to some early reports (Dieminger et al, 1976; Volanakis et al, 1977) and suggested that the factor $\bar{D}$ preparations used in earlier studies may have contained other contaminating proteases.

Although ZLSB was apparently slowly hydrolysed by factor $\bar{D}$, the hydrolysis rates in three separate assays monitored over half-hour intervals varied from 0-0.3 µmoles/min/mg. Fractions from a CM-cellulose 32 column taken from the 18-22 mS region of the gradient had no ability to hydrolyse ZLSB, nor did the 25-28 mS glycoprotein eluting immediately after factor $\bar{D}$.

5.4.3 p-NPGB

When p-NPGB binds to the active site of a serine protease, p-nitrophenol is released and can be quantitated by measuring ΔA_{402}. Unlike the substrates discussed above, release of

Table 5.4 Esterolytic activity of factor $\bar{D}$ and other serine proteases

(a) Arginyl esters

Enzyme	Substrate hydrolysis (μmole/min/mg) using TAME (1 mM)	BAEE (0.5 mM)
Trypsin	587	53.3
Chymotrypsin(1)	0.005	0.030
$C\bar{1}r$(2,3)	0	0
$C\bar{1}s$(2)	9.53	0
Factor $\bar{D}$	0	0

Substrate hydrolysis was measured spectrophotometrically at 247 nm (TAME) or 253 nm (BAEE). Assays were carried out in duplicate or triplicate as described in section 2.2.6. Enzyme concentrations used were: trypsin, 1.8×10^{-9} M; chymotrypsin, 1.8×10^{-6} M; $C\bar{1}r$, 1×10^{-7} M; $C\bar{1}s$, 3.2×10^{-8} M; factor $\bar{D}$, 7.6×10^{-8} M.

(1) Chymotrypsin was treated with SBTI to inhibit residual trypsin activity.

(2) Andrews and Baillie (1979), using 0.05 mM TAME and 10 mM BAEE.

(3) Derived from Volanakis et al (1977) using 10 mM TAME and 10 mM BAEE.

(b) ZLSB (10 mM)

Enzyme	Substrate hydrolysis (μmol/min/mg)
Trypsin	58.6 ± 1.1
Chymotrypsin	0.069 ± 0.069
$C\bar{1}r$(1)	0.101
$C\bar{1}s$(1)	0.760
Factor $\bar{D}$	0.015 ± 0.015

Substrate hydrolysis was measured spectrophotometrically at 412 nm. Assays were carried out in duplicate or triplicate as described in section 2.2.6. Enzyme concentrations used were: trypsin, 1.4×10^{-9} M; chymotrypsin (treated with SBTI), 1.4×10^{-7} M; $C\bar{1}r$, 1×10^{-8} M; $C\bar{1}s$, 5×10^{-9} M; factor $\bar{D}$, 4.2×10^{-8} M.

(1) A.W. Dodds (unpublished results, using 1 mM ZLSB).

nitrophenol occurs in a single burst, after which the turnover of p-NPGB is slow and depends on the rate of decay of the acyl-enzyme. By extrapolating the post-burst A_{402} to zero-time and then measuring ΔA_{402} at this point, the amount of p-nitrophenol released can be quantitated.

The cleavage of p-NPGB by factor $\bar{D}$ derived from a CM-cellulose 32 column, is given in Table 5.5. The results require confirmation by assaying across the column using maximal amounts of enzyme, but they do suggest that p-NPGB is hydrolysed by factor $\bar{D}$. The specific activity (μmoles p-nitrophenol released/mg enzyme) of factor $\bar{D}$ towards p-NPGB was about 5-10 times that of the 25-28 mS crude glycoprotein, whereas the 18-22 mS protein was devoid of activity. The amount of p-nitrophenol released in the burst is equivalent to the concentration of active enzyme, so that if all the enzyme present is active, the specific activity (expressed as μmoles p-nitrophenol released/μmole enzyme) would be 1.0. The value obtained for factor $\bar{D}$ was 1.7 ± 0.8 (compared to 1.13 ± 0.04 for trypsin). The ΔA_{402} following the initial burst of p-nitrophenol release by factor $\bar{D}$ was zero, suggesting that decay of the acyl-enzyme was slow.

5.4.4 Conclusions

The results presented here support the observations of others (Lesavre and Müller-Eberhard, 1978; Davis *et al*, 1979b) that factor $\bar{D}$ shows little activity toward synthetic substrates. The observed activity ranged from 0.0008-0.06% that observed with trypsin, except for p-NPGB which was hydrolysed with comparable efficiency by both enzymes.

Table 5.5 Cleavage of p-NPGB by factor $\bar{D}$ and other fractions from CM-cellulose 32

Enzyme	p-nitrophenol released (10^3 x μmoles/mg)
Trypsin	48.7 ± 1.3
18-22 mS protein	0
Factor $\bar{D}$	72.8 ± 36.4
25-28 mS protein	7.2 ± 1.8

Release of p-nitrophenol was measured spectrophotometrically at 402 nm. Details of the assay are described in the text. Enzyme concentrations used were: trypsin, 2.0×10^{-6} M; 18-22 mS protein, approx. 3.3×10^{-7} M (calculated assuming an apparent mol. wt. of 12,000 and $E^{1\%}_{280} = 10$); factor $\bar{D}$, 0.75×10^{-7} M; 25-28 mS protein, approx. 2.7×10^{-6} M (calculated assuming apparent mol. wt. of 24,000 and $E^{1\%}_{280} = 15$).

Factor $\bar{D}$ was less active that either $C\bar{1}r$ or $C\bar{1}s$ towards nitroanilide and lysyl substrates except for the nitroanilide substrate, S2238 (Table 5.3), and was slightly more active than SBTI-treated chymotrypsin. In all cases, it could not be excluded that the small amount of activity observed was caused by a contaminating protease or was due to experimental error. However, the hydrolysis of p-NPGB by factor $\bar{D}$ appeared to be authentic, although because of the high experimental error observed, confirmation will be necessary.

5.5 INHIBITOR STUDIES

The ability of a variety of protease inhibitors including some plasma protease inhibitors, to inhibit factor $\bar{D}$ haemolytic activity was investigated. The inhibitors together with the conditions of pre-incubation with factor $\bar{D}$ are listed in Table 2.2. Following pre-incubation, haemolytic assays were carried out immediately in duplicate together with controls consisting of inhibitor plus buffer. Except where stated otherwise, crude factor $\bar{D}$ from Sephadex G-75 fractionation of normal human serum was used.

Of all the inhibitors tested, only iPr_2P-F and p-NPGB had any effect on factor $\bar{D}$ haemolytic activity (Tables 5.6 and 5.7). Complete inhibition of haemolytic activity was observed with iPr_2P-F at a concentration of >2 mM. At 1-2 mM, the haemolytic activity was substantially reduced, although the amount of residual activity varied depending on the source of the factor $\bar{D}$ and whether inhibition was followed by a dialysis step (Table 5.6(b)). Samples of crude factor $\bar{D}$

Table 5.6 Inhibition of factor $\bar{D}$ by iPr_2P-F

(a)

iPr_2P-F (mM)	Haemolytic activity (units/10 µl)
0	1.6
1	<0.01
2	0
5	0
10	0

Crude factor $\bar{D}$ (approx. 3 µg/ml) was inhibited as described in section 2.2.7. Samples were stood for 20 h at 4°C before assaying in duplicate. Isopropanol alone had no effect on haemolytic activity.

(b) Residual activity following inhibition with 2 mM iPr_2P-F

Source of crude factor $\bar{D}$	Residual activity (units/10 µl) after inhibition	
	Dialysed	Undialysed
QAE-Sephadex A-50		
"precursor factor $\bar{D}$"	0.043	0
factor $\bar{D}$	0.024	0.013
Sephadex G-75	0	0

Factor $\bar{D}$ was inhibited with 2 mM iPr_2P-F as described in section 2.2.7. Samples were then dialysed for 20 h at 4°C or else stood for the same period at 4°C, before assaying in duplicate.

Table 5.7 Inhibition of factor $\bar{D}$ by p-NPGB

p-NPGB (mM)	Haemolytic activity (units/10 µl)
0	0.40
1	0.40
2	0.24
5	0
10	0

Factor $\bar{D}$ was inhibited with p-NPGB as described in section 2.2.7. Dimethylformamide alone had no effect on activity. Activity was not regained by either (i) standing at 4°C for 24 h to allow hydrolysis of excess p-NPGB or by (ii) dialysis for 24 h at 4°C.

in 0.0035 M sodium phosphate, pH 8.0 from QAE-Sephadex A-50 (section 2.2.4.1) which were dialysed for 20 h at 4°C following the inhibition step, retained about twice as much haemolytic activity as undialysed samples which had stood at 4°C for the same time period. On the other hand, crude factor $\bar{D}$ in 0.1 M Tris/0.2 M NaCl/2 mM EDTA, pH 8.0 obtained from the 25,000 dalton region of a Sephadex G-75 column was completely inhibited by 2 mM iPr_2P-F, and the dialysis step had no effect.

Inhibition by p-NPGB was observed at a concentration of 2 mM (Table 5.7) using both crude factor $\bar{D}$ and factor $\bar{D}$ from a CM-cellulose 32 column (section 2.2.4.3). The results were unaltered by standing the inhibited samples for 24 h at 4°C to allow hydrolysis of excess pNPGB or after dialysis for 24 h at 4°C.

None of the other inhibitors tested, including the plasma protease inhibitors (Table 2.2) were effective in inhibiting factor $\bar{D}$ haemolytic activity which was retained within 90-100% of the control. The lack of inhibition by non-serine protease inhibitors was consistent with the serine protease nature of factor $\bar{D}$ while the lack of inhibition by common plasma protease inhibitors supported the observation of Lesavre and Müller-Eberhard (1978) that factor $\bar{D}$ specific activity (haemolytic activity/ µg factor $\bar{D}$ antigen) does not increase during purification. It was not entirely unexpected that synthetic substrates had no effect on haemolytic activity: because they behave as competitive inhibitors, high substrate concentrations

are normally required to cause effective inhibition. However, it was hoped that evidence might be obtained as to whether factor $\bar{D}$ bound the substrates, even although the spectrophotometric assays showed that no significant hydrolysis occurred (section 5.4). Although this point was not clarified, the results of others suggest that certain synthetic substrates will bind to the active site: by preincubation of factors B, $\bar{D}$ and CVF with 26 mM ALME or 20 mM AGLME, Dieminger *et al* (1976) found that the generation of C3-cleaving activity was reduced by 50%; Fearon and Austen (1975b) have observed competitive inhibition by 10 mM TAME; and Lesavre *et al* (1980) have shown that peptides resembling the cleavage site in factor B may also cause inhibition of factor $\bar{D}$ activity. In view of these observations it is possible that the lack of inhibition observed in the experiments reported here was a function of the assay procedure used.

CHAPTER 6

CONCLUSION

6.1 A COMPARISON OF THE PRIMARY STRUCTURE OF FACTOR $\bar{D}$ WITH OTHER SERINE PROTEASES

From the sequence data on factor $\bar{D}$ (Fig. 4.33) it can be seen that of the amino acid residues obtained, virtually all the residues invariant in other serine proteases were also found in equivalent positions in factor $\bar{D}$. An exception was CTG residue 184 (usually glycine) which was replaced by Glu in factor $\bar{D}$. Asp_{194} which forms an ion pair with Ile_1 was present, as were the active site residues, Asp_{102} and Ser_{195} which are part of the charge-relay system. $Residue_{57}$ which would correspond to the active site histidine was not identified, but the conservation in factor $\bar{D}$ of the sequence on either side of the missing residue, suggests that this will be found to be His.

Of residues which are highly conserved in serine proteases, many were also found in factor $\bar{D}$. The specificity residue 189 was Asp as was expected in an arginine-specific serine protease. Residue 99 with which the P_2 sidechain probably interacts in other serine proteases (Segal, 1971) was threonine. Residue 40 which in pancreatic serine proteases is His, is thought to be involved in the stabilisation of the zymogen (Freer et al, 1970; de Haën et al, 1975). It does not appear to be highly conserved in non-pancreatic serine proteases, and was not present in factor $\bar{D}$ in the results reported here. However, Volanakis et al (1980) have identified His in this position. An interesting

substitution was found at residue 192 which in other serine proteases is deeply buried in the zymogen. On activation, it shifts to the surface of the enzyme and forms a flexible "lid" to the binding pocket (section 1.3.1.2). Serine proteases with an Arg/Lys specificity tend to have Gln in this position (although thrombin and factor B are exceptions) whereas Lys was found in factor $\bar{D}$.

In Fig. 6.1 specific regions of the amino acid sequence of factor $\bar{D}$ have been compared with other mammalian proteases including some complement proteases for which published data are available. A preliminary statistical evaluation of the differences between amino acid sequences can be obtained using the method of Moore and Goodman (1977) which is based on the minimum number of base changes required to make two sequences identical over a given number of residues. Using this method, the N-terminal sequence of factor $\bar{D}$ was compared with other serine proteases as far as residue 35 in Fig. 6.1(a) (data beyond this point were not included as no data were available for $C\bar{1}r$ and $C\bar{1}s$ b chains). It is clear from this comparison that the N-terminal sequence is more closely related to those of rat group specific protease (GSP) and porcine elastase than to the equivalent sequences in other plasma serine proteases (Table 6.1). In this respect it is of interest that the arrangement of disulphide bonds suggested by the positioning of half-cystine residues, is also similar to elastase (Fig. 1.5) i.e. (CTG numbering) Cys 42-58 (the "His" loop); Cys 168·182 (the "Met" loop); probably Cys 191-220 (the "Ser" loop. Residue 220 which has not yet been identified, is invariant in other serine proteases); and Cys 136-201. The similarity between the

Fig. 6.1 Comparison of N-terminal and active site regions of factor $\bar{D}$ with the equivalent regions in other serine proteases

Sequence data are from Sim et al, 1977 ($C\bar{1}r$ and $C\bar{1}s$); Woodbury et al, 1978 (group specific protease); Katayama et al, 1979 (Factor IX); Christie et al, 1980 (Factor B), de Haën et al (1975) and Dayhoff, 1979 (elastase, chymotrypsin, trypsin, thrombin, plasmin and factor X). The residue numbering is that of the CTG sequence (Dayhoff, 1979). The active site residues (serine, histidine and aspartic acid) are underlined in each case, except for the active site histidine of factor $\bar{D}$ which was not obtained in the studies reported here.

Protein	16 . . . 20 . . . . 30 . . . . 40 . . . . 50 58
Factor $\bar{D}$	I L G G R E A E A H A R P Y M A S V Q L - - - - N G A E L C G G V L V A E Q W V L T A A X C
Group specific protease (rat)	I I G G V E S I P H S R P Y M A H L D I V T E K G L R V I C G G F L I S R Q F V L T A A H C
Elastase (pig)	V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V M T A A H C
C$\bar{1}$r b chain (human)	I I G G Q K A K M G N F P W Q V F T N Z
C$\bar{1}$s b chain (human)	I I G G S D A D I K N F P W Q V F F D N
Chymotrypsin A (cow)	I V N G E E A V P G S W P W Q V S L Q D K T G F - - - H F C G G S L I N E N W V V T A A H C
Plasmin (human)	V V G G C V A H P H S W P W Q V S L R T R F G M - - - H F C G G T L I S P E W V L T A A H C
Trypsin (cow)	I V G G Y T C G A N T V P Y Q V S L - - N S G Y - - - H F C G G S L I N S Q W V V S A A H C
Thrombin (cow)	I V E G Q D A E V G L S P W Q V M L F R K S P Q E - - L L C G A S L I S D R W V L T A A H C
Factor X (cow)	I V G G R D C A E G E C P W Q A L L V - N E E N E - - G F C G G T I L N E F Y V L T A A H C
Factor IX (cow)	V V G G E D A E R G Q F P W Q V L L H G - - - - E I A A F C G G S I V N E K W V V T A A H C
Factor B b chain (human)	W E H R K G T D Y H K Q P W Q A K I S V I R P S K G X E S C M G A V V S E Y F V L T A A H C

Fig. 6.1(a) The N-terminal region and the region of the active site histidine

Continued......

	102									112
Factor D̄	D	L	L	L	L	Q	L	S	E	K	A
Group specific protease (rat)	D	I	M	L	L	K	L	E	K	K	V
Elastase (pig)	D	I	A	L	L	R	L	A	Q	S	V
Chymotrypsin A (cow)	D	I	T	L	L	K	L	S	T	A	A
Plasmin (human)	D	I	A	L	L	K	L	S	S	P	A
Trypsin (cow)	D	I	M	L	I	K	L	K	S	A	A
Thrombin (cow)	D	I	A	L	L	K	L	K	R	P	I
Factor X (cow)	D	I	A	V	L	R	L	K	T	P	I
Factor IX (cow)	D	I	A	L	L	E	L	D	E	P	L
Factor B b chain (human)	D	V	A	L	I	K	L	K	N	K	L

Fig. 6.1(b) The region of the active site aspartic acid

	180															190										200
Factor D̄	M	-	C	A	E	S	-	-	-	-	N	R	R	-	D	S	C	K	G	D	S	G	G	P	L	V	C
Group Specific protease (rat	Q	V	C	V	G	S	-	-	P	T	T	L	R	-	A	A	F	M	G	D	S	G	G	P	L	L	C
Elastase (pig)	M	V	C	A	G	-	-	-	G	N	G	V	R	-	S	G	C	Q	G	D	S	G	G	P	L	H	C
Chymotrypsin A (cow)	M	I	C	A	G	-	-	-	A	S	G	V	-	-	S	S	C	M	G	D	S	G	G	P	L	V	C
Plasmin (human)	E	L	C	A	G	H	-	-	L	A	G	G	T	-	D	S	C	Q	G	D	S	G	G	P	L	V	C
Trypsin (cow)	M	F	C	A	G	Y	-	-	L	E	G	G	K	-	D	S	C	Q	G	D	S	G	G	P	V	V	C
Thrombin (cow)	M	F	C	A	G	Y	K	P	G	E	G	K	R	G	D	A	C	E	G	D	S	G	G	P	F	V	M
Factor X (cow)	M	F	C	A	G	Y	-	P	D	T	Q	P	E	-	D	A	C	Q	G	D	S	G	G	P	H	V	T
Factor IX (cow)	M	F	C	A	G	Y	-	-	H	E	G	G	K	-	D	S	C	Q	G	D	S	G	G	P	H	V	T
Factor B b chain (human)	F	L	C	T	G	G	V	S	P	Y	A	D	P	-	N	T	C	R	G	D	S	G	G	P	L	I	V

Fig. 6.1(c) The region of the active site serine

Table 6.1 Comparison of factor $\bar{D}$ (residues 1-20) with other serine proteases

Protease	Probability of a random relationship between sequences
Group specific protease (Rat)	<0.01%
Elastase (Pig)	0.01%
$C\bar{1}r$ b chain (Human)	5.0%
$C\bar{1}s$ b chain (Human)	5.0%
Chymotrypsin A (Cow)	0.1%
Plasmin (Human)	0.1%
Trypsin (Cow)	0.1%
Thrombin (Cow)	>1.0%<5.0%
Factor X (Cow)	>0.1%<1.0%
Factor IX (Cow)	10%
Factor B b chain (Human)	>10%<50%

The sequence of factor $\bar{D}$ from residues 1-20 was compared with the equivalent sequence in each of the other proteases using the method of Moore and Goodman (1977). The source of the data was the same as for Fig. 6.1.

N-terminal sequences of factor $\bar{D}$ and GSP is also interesting as neither protease appears to have a zymogen.

The sequences about the active site histidine (residues 42-58, Fig. 6.1a) and the active site serine (residues 180-201, Fig. 6.1c) show extremely high homology, with the probability of a random relationship between factor $\bar{D}$ and other sequences $<<0.01\%$, except for the sequence around the active site serine of GSP for which the figure is 0.01-0.1%. The statistic for the region of the active site aspartic acid (residues 102-112, Fig. 6.1b) is more variable, with the probability of a random relationship ranging from 0.01% (GSP) to 1% except for Bb (5%), thrombin (10%) and factor X (10-50%).

From the overall similarity to other serine proteases, it is probable that factor $\bar{D}$ will have a similar gross topography to other vertebrate serine proteases, although it seems likely that it will be more closely related to enzymes of pancreatic origin than of plasma origin. The changes in some residues which by analogy, occur in the region of the active site, may help explain the unusual specificity of this enzyme.

6.2 THE SPECIFICITY AND REGULATION OF FACTOR $\bar{D}$ ACTIVITY

6.2.1 Specificity

The present knowledge of factor $\bar{D}$ suggests that

(1) no detectable zymogen occurs

(2) the catalytic site in factor $\bar{D}$ is accessible to small inhibitors such as iPr_2P-F, cyclohexylbutyl-phosphonofluoridate and NPGB.

(3) the binding site is available to small molecules such as certain synthetic substrates (Fearon and Austen, 1975; Dieminger et al, 1976) and probably to the inhibitors diamino-propane and diamino-butane (K.J. Gadd, unpublished results).

(4) substrates which mimic the cleavage site in factor B and contain the P_1-P_3 and P'_1-P'_3 residues[†] will bind to factor $\bar{D}$ but will not be cleaved (Lesavre et al, 1980).

From these observations it can be concluded that the active site of factor $\bar{D}$ is essentially preformed in the free enzyme and substrate binding is possible, but cleavage does not occur, even of peptides which resemble the cleavage site in factor B. No evidence is available to explain this enigma, but several factors can contribute to the reactivity of an enzyme toward a given substrate. These include the extent of substrate binding, the "productivity" of the binding mode and the rates of acylation and deacylation.

Some proteases show extensive subsite specificity at the active site, extending over perhaps seven or eight residues

[†]P and P' residues of a substrate, and S and S' residues of an enzyme, are defined in section 1.3.1.2.

(Schechter and Berger, 1967; Abramovitz et al, 1967; Atlas et al, 1970; Atlas, 1975). In the case of elastase, important enzyme-substrate interactions occur as far as S_4 and S_5, and S' interactions are also important in enhancing acylation (Thompson and Blout, 1970). Binding of a substrate or inhibitor to serine proteases is accompanied by reversible conformational changes which may vary using different substrates or inhibitors, and which are stabilised by formation of the acyl-enzyme intermediate (Krieger et al, 1974; Moon et al, 1965a, Kraut, 1977). Furthermore, enzymes such as CT and trypsin and their zymogens can adopt "active" or "inactive" conformations (Oppenheimer et al, 1966). At neutral pH, about 15% of α-CT is in the inactive conformation (Ferscht, 1972). The active conformation is stabilised not only by protonation of the N-terminal isoleucine, but also by substrate binding (Hess, 1971; Moon et al, 1965a; 1965b). The questions can then be raised as to

(a) whether factor $\bar{D}$ has an extensive binding site, so that small substrates will be bound but in a non-productive mode because interactions remote from S_1 are important.

(b) whether factor $\bar{D}$ may circulate in a largely inactive conformation and that only the "correct" substrate will induce stabilisation of an active conformation.

The sequence at the susceptible site in factor B is

-lys-arg-lys

P_2 P_1 ↑ P'_1

and the specificity of factor $\bar{D}$ must be related in part to the need to accommodate this arrangement of three positively charged residues. However, the tertiary structure of the cleavage site in C3b-bound factor B must also be important as in free factor B the site is accessible to trypsin but not to factor $\bar{D}$ (section 1.4.3).

The reason for the substrate specificity of factor $\bar{D}$ may be clearer following the complete elucidation of its amino acid sequence. Kinetic studies with substrates similar to those used by Lesavre et al (1980) may be of use, but it is possible that X-ray crystallographic data will be necessary to give a completely satisfactory explanation.

6.2.2 Regulation

A physiological activator of factor $\bar{D}$ has not been described and at present, there is no evidence that a zymogen form of factor $\bar{D}$ exists. Furthermore, it appears to completely resist the regulatory activity of the plasma protease inhibitors, including α_2-macroglobulin (section 5.5). However, its remarkable substrate specificity in itself acts as a form of regulation, preventing uncontrolled activation of free factor B. Secondly, the activity in vivo is also controlled by the availability of nascent C3b to which factor B must bind in order to become activated. The importance of this latter mechanism is exemplified in the case of C3b INA deficiency where, not only does C3 depletion occur, but factor B levels are also markedly reduced (Alper and Rosen, 1971).

REFERENCES

ABRAHAM, G.N., GOLDBERGER, G.G., COLTEN, H.R. WILLIAMS, J. (1980) Fed. Proc. Amer. Soc. Exp. Biol., 39, 1201 (abstract)

ABRAMOVITZ, N., SCHECHTER, I., BERGER, A. (1967) Biochem. Biophys. Res. Commun., 29, 862.

ALLEN, A.K., NEUBERGER, A. (1975) FEBS Letters, 60, 76.

ALPER, C.A., ROSEN, F.S. (1971) Adv. Immunol., 14, 251.

ALPER, C.A., ROSEN, F.S., LACHMANN, P.J. (1972) Proc. Natl. Acad. Sci. U.S.A., 69, 2910.

ANDREWS, J.M., BAILLIE, R.D. (1979) J. Immunol., 123, 1403.

ATLAS, D. (1975) J. Mol. Biol., 93, 39.

ATLAS, D., LEVIT, S., SCHECHTER, I., BERGER, A. (1970) FEBS Letters, 11, 281.

AUGENER, W., GREY, H.M., COOPER, N.R., MÜLLER-EBERHARD, H.J. (1971) Immunochem., 8, 1011.

BARNSTABLE, C.J. (1978) TIBS, 3, N158.

BARRETT, A.J., STARKEY, P.M. (1973) Biochem. J., 133, 709.

BENDER, M.L., KILLHEFFER, J.V. (1973) CRC Crit. Rev. Biochem., 1, 149.

BENTLEY, C., BITTER-SUERMANN, D., HADDING, U., BRADE, V. (1976) Eur. J. Immunol., 6, 393.

BENTLEY, C., FRIES, W., BRADE, V. (1978) Immunol., 35, 971.

BENTLEY, C., HADDING, U., BITTER-SUERMANN, D., BRADE, V. (1977) Eur. J. Immunol., 7, 188.

BHAKDI, S., TRANUM-JENSEN, J. (1978) Proc. Natl. Acad. Sci. U.S.A., 75, 5655.

BIRKTOFT, J.J., BLOW, D.M., HENDERSON, R., STEITZ, T.A. (1970) Phil. Trans. Roy. Soc. London (1970) B257, 67.

BLAKE, C.C.F. (1979) in "Companion to Biochemistry", 2, 281 (ed. A.T. Bull, J.R. Lagnado, J.O. Thomas, K.F. Kipton) Longmans, London.

BLOW, D.M. (1971) in "The Enzymes", 3, 185 (ed. P.D. Boyer) Acad. Press, New York.

BLOW, D.M., BIRKTOFT, J.J., HARTLEY, B.S. (1969) Nature, 221, 337.

BODE, W., HUBER, R. (1976) FEBS Letters, 68, 231.

BODMER, W.F. (1976) Harvey Lectures, 72, 91.

BOOTH, N.A., CAMPBELL, R.D., SMITH, M.A., FOTHERGILL, J.E. (1979) Biochem. J., 183, 573.

BORDET, J. (1896) Ann. Inst. Pasteur Paris, 10, 193.

BOYLE, M.D.P., BORSOS, T. (1980) Mol. Immunol., 17, 425.

BRADE, V., BENTLEY, C., BITTER-SUERMANN, D., HADDING, U. (1977a) Z. Immun-Forsch., 152, 402.

BRADE, V., DIEMINGER, L., SCHMIDT, G., VOGT, W. (1976) Immunol., 30, 171.

BRADE, V., HALL, R.E., COLTEN, H.R. (1977b) J. Exp. Med., 146, 759.

BRADE, V., NICHOLSON, A., BITTER-SUERMANN, D., HADDING, U. (1974b) J. Immunol., 113, 1735.

BRADE, V., NICHOLSON, A., LEE, G.D., MAYER, M.M. (1974a) J. Immunol., 112, 1845.

BRAUN, H. (1911) Biochem. Z., 31, 65.

BRIDGEN, P.J., CROSS, G.A., BRIDGEN, J. (1976) Nature, 263, 613.

BRODSKY-DOYLE, B., LEONARD, K.R., REID, K.B.M. (1976) Biochem. J., 159, 279.

BURGER, R., BITTER-SUERMANN, D., HADDING, U. (1978) Immunochem., 15, 231.

BURGER, R., HADDING, U., SCHORLEMMER, H.U., BRADE, V., BITTER-SUERMANN, D. (1975) Immunol., 29, 549.

BUCHNER, H. (1889) Centralbl. Backteriol. Parasitol., 5, 817.

CAMPBELL, R.D., DODDS, A.W., PORTER, R.R. (1980) Biochem. J., 189, 67.

CAPEL, P.J.A., GROENEBOER, O., GROSVELD, G., PONDMAN, K.W. (1978) J. Immunol., 121, 2566.

de CEULAER, C., PAPAZOGLOU, S., WHALEY, K. (1980) Immunol., 41, 37.

CHASE, M.W., WILLIAMS, C.A. (1968) in "Methods in Immunology and Immunochemistry", 2, 251 (ed. C.A. Williams and M.W. Chase) Acad. Press, New York.

CHRISTIE, D.L., GAGNON, J., PORTER, R.R. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 4923.

COLOMB, M., PORTER, R.R. (1975) Biochem. J., 145, 177.

COLTEN, H.R. (1980) Progr. Immunol., 4 (in press).

COOPER, N.R. (1971) Progr. Immunol., 1, 567.

COOPER, N.R. (1975) Biochem., 14, 4245.

COOPER, N.R., MÜLLER-EBERHARD, H.J. (1968) Immunochem., 5, 155.

COOPER, N.R., MÜLLER-EBERHARD, H.J. (1970) J. Exp. Med., 132, 775.

CUNNINGHAM, C.M., KINGZETTE, M., RICHARDS, R.L., ALVING, C.R., LINT, T.F.. GEWURZ, H. (1979) J. Immunol., 122, 1237.

DAHA, M.R., AUSTEN, K.F., FEARON, D.T. (1978) J. Immunol., 120, 1389.

DAHA, M.R., FEARON, D.T., AUSTEN, K.F. (1976) J. Immunol., 116, 568.

DAHA, M.R., HAZEVOET, H.M., van ES, L.A., CATS, A. (1980) Immunol., 40, 417.

DAVIS, A.E. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 4938.

DAVIS, A.E., KENNEY, D.M. (1979) J. Clin. Invest., 64, 721.

DAVIS, A.E., ZALUT, C., ROSEN, F.S., ALPER, C.A. (1979a) Fed. Proc. Fed. Am. Soc. Exp. Biol., 38, 1216 (abstract).

DAVIS, A.E., ZALUT, C., ROSEN, F.S., ALPER, C.A. (1979b) Biochem., 18, 5082.

DAVIS, A.E., ZIEGLER, J.B., GELFAND, E.W., ROSEN, F.S., ALPER, C.A. (1977) Proc. Natl. Acad. Sci. U.S.A., 74, 3980.

DAWSON, R.M.C., ELLIOTT, D.C., ELLIOTT, W.H., JONES, K.M. (1978) "Data for Biochemical Research" p.476 (O.U.P., Oxford).

DAY, N.K.,SCHREIBER, R.D., GÖTZE, O., MÜLLER-EBERHARD, H.J. (1976) Scand. J. Immunol., 5, 715.

DAYHOFF, M.O. (1979) "Atlas of Protein Sequence and Structure" Vol. 5, Supplement 3, 1978, p.73 (National Biomedical Research Foundation, Washington).

DIAS da SILVA, W., LEPOW, I.H. (1967) J. Exp. Med., 125, 921.

DIEMINGER, L., VOGT, W., LYNEN, R. (1976) Z. Immun-Forsch., 152, 231.

DIFRICH, M.P., HADDING, U., KÖNIG, W., LIMBERT, M., SCHORLEMMER, H-U, BITTER-SUERMANN, D. (1974) Immunochem., 11, 527.

DLOUHA, V., KEIL, B. (1969) FEBS Letters, 3, 137.

DODDS, A.W., PORTER, R.R. (1979) Mol. Immunol., 16, 1059.

DODDS, A.W., SIM, R.B., PORTER, R.R., KERR, M.A. (1978) Biochem. J., 175, 383.

DOURMASHKIN, R.R. (1978) Immunol., 35, 205.

DRAPEAU, G.R., BOILY, Y., HOUMARD, J. (1972) J. Biol. Chem., 247, 6720.

von DUNGERN, E. (1900) Münch. Med. Wochenchr. 47, 677.

EDMAN, P., BEGG, G. (1967) Eur. J. Biochem., 1, 80.

EDSTROM, R.D. (1968) Anal. Biochem., 26, 204.

EDWARDS, M.S., NICHOLSON-WELLER, A., BAKER, C.J., KASPER, D.L. (1980) J. Exp. Med., 151, 1275.

ESSER, A.F., KOLB, W.P., PODACK, E.R., MÜLLER-EBERHARD, H.J. (1979) Proc. Natl. Acad. Sci. U.S.A., 76, 1410.

FAIRBANKS, G., STECK, T.L., WALLACH, D.F.H. (1971) Biochem., 10, 2606.

FEARON, D.T. (1978) Proc. Natl. Acad. Sci. U.S.A., 75, 1971.

FEARON, D.T., AUSTEN, K.F. (1975a) Proc. Natl. Acad. Sci. U.S.A., 72, 3220.

FEARON, D.T., AUSTEN, K.F. (1975b) Ann. N.Y. Acad. Sci., 256, 441.

FEARON, D.T., AUSTEN, K.F. (1976) Essays Med. Biochem., 2, 1.

FEARON, D.T., AUSTEN, K.F. (1977a) Proc. Natl. Acad. Sci. U.S.A., 74, 1683.

FEARON, D.T., AUSTEN, K.F. (1977b) J. Exp. Med., 146, 22.

FEARON, D.T., AUSTEN, K.F., RUDDY, S. (1973) J. Exp. Med., 138, 1305.

FEARON, D.T., AUSTEN, K.F., RUDDY, S. (1974) J. Exp. Med., 139, 355.

FERSCHT, A.R. (1972) J. Mol. Biol., 64, 497.

von FODOR, J. (1886) Deut. Med. Wochenschr., 13, 745.

FOTHERGILL, J.E., ANDERSON, W.H.K. (1978) Curr.Topics Cell Regul., 13, 259.

FREER, S.T, KRAUT, J., ROBERTUS, J.D., WRIGHT, H.T., XUONG, N-H. (1970) Biochem., 9, 1997.

FUJITA, T., NUSSENZWEIG, V. (1979) J. Exp. Med., 150, 267.

FUJITA, T., TAKIUCHI, M., IIDA, K., NAGAKI, K., INAI, S. (1977) Immunochem., 14, 25.

GADD, K.J., REID, K.B.M. (1981a) Immunol. (in press).

GADD, K.J., REID, K.B.M. (1981b) Biochem. J. (submitted for publication).

GARDNER, W.D., HASLEBY, J.A., HOCH, S.O. (1980) J. Immunol., 124, 2800.

GERTLER, A., WALSH, K.A., NEURATH, H. (1974) Biochem., 13, 1302.

GIAVEDONI, E.B., DALMASSO, A.P. (1976) J. Immunol., 116, 1163.

GITLIN, J.D., ROSEN, F.S., LACHMANN, P.J. (1975) J. Exp. Med., 141, 1221.

GIGLI, I., AUSTEN, K.F. (1969a) J. Exp. Med., 129, 679.

GIGLI, I., AUSTEN, K.F. (1979b) J. Exp. Med., 130, 833.

GOERS, J.W.F., PORTER, R.R. (1978) Biochem. J., 175, 765.

GOETZL, E.J., AUSTEN, K.F. (1974) in "Chemotaxis: its Biology and Biochemistry" p.218 (ed. E. Sorkin and Davos-Platz) S. Karger, Basel.

GOLDBERGER, G., COLTEN, H.R. (1980) Nature, 286, 514.

GOLDLUST, M.B., SHIN, H.S., HAMMER, C.H., MAYER, M.M.(1974) J. Immunol., 113, 998.

GORSKI, J.P., HUGLI, T.E., MÜLLER-EBERHARD, H.J. (1979a) Fed. Proc. Fed. Am. Soc. Exp. Biol., 38, 1010.

GORSKI, J.P., HUGLI, T.E., MÜLLER-EBERHARD, H.J. (1979b) Proc. Natl. Acad. Sci., U.S.A., 76, 5299.

GÖTZE, O. (1976) Fed. Proc. Fed. Am. Soc. Exp. Biol., 35, 254.

GÖTZE, O., MÜLLER-EBERHARD, H.J. (1971) J. Exp. Med., 134, 90s.

GÖTZE, O., MÜLLER-EBERHARD, H.J. (1976) Adv. Immunol., 24, 1.

GOUNARIS, A.D., PERLMANN, G.E. (1967) J. Biol. Chem., 242, 2739.

GREEN, G.D.J., SHAW, E. (1979) Anal. Biochem., 93, 223.

GREFRATH, S.P., REYNOLDS, J.A. (1974) Proc. Natl. Acad. Sci. U.S.A., 71, 3913.

GROSS, E. (1967) Methods in Enzymology, 11, 238.

GROSS, E., WITKOP, B. (1962) J. Biol. Chem., 237, 1856.

de HAËN, C., NEURATH, H., TELLER, D.C. (1975) J. Mol. Biol., 92, 225.

HADDING, L., HAMURO, J., BITTER-SUERMANN, D. (1978) J. Immunol., 120, 1776 (abstract).

HALL, R.E., COLTEN, H.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 76, 1707.

HAMMER, C.H., ABRAMOVITZ, A.S., MAYER, M.M. (1976) J. Immunol., 117, 830.

HAMMOND, B.R., GUTFREUND, H. (1955) Biochem. J., 61, 187.

HARPEL, P.C., COOPER, N.R. (1975) J. Clin. Invest., 55, 593.

HARTLEY, B.S. (1970) Biochem. J., 119, 805.

HARTLEY, B.S., SHOTTON, D.M. (1971) in "The Enzymes", 3, 323 (ed. P.D. Boyer) Acad. Press, New York.

HAUPT, H., HEIMBURGER, N., KRANZ, T., SCHWICK, H.G. (1970) Eur. J. Biochem., 17, 254.

HAYASHI, R. (1977) Methods in Enzymology, 47, 84.

HENDERSON, R. (1970) J. Mol. Biol., 54, 341.

HESS, G.P. (1971) in "The Enzymes" 3, 213 (ed. P.D. Boyer) Acad. Press, New York.

HOBART, M.J., LACHMANN, P.J. (1976) Transplant. Rev., 32, 26.

HOUMARD, J., DRAPEAU, G.R. (1972) Proc. Natl. Acad. Sci. U.S.A., 69, 3506.

HUGHES-JONES, N.C., GARDNER, B. (1979) Mol. Immunol., 16, 697.

HUGHES-JONES, N.C., GARDNER, B., ROWLANDS, J. (1977) Nature, 270, 613.

HUNKAPILLER, M.W., HOOD, L.E. (1978) Biochem., 17, 2124.

HUNKAPILLER, M.W., SMALLCOMBE, S.H., WHITACKER, D.R., RICHARDS, J.H. (1973) Biochem., 12, 4732.

HUNSICKER, L.G., RUDDY, S., AUSTEN, K.F. (1973) J. Immunol., 110, 128.

HURST, M.M., VOLANAKIS, J.E., STROUD, R.M., BENNETT, J.C. (1975) J. Exp. Med., 142, 1322.

ISENMAN, D.E., DORRINGTON, K.J., PAINTER, R.H. (1975) J. Immunol., 114, 1726.

ISENMAN, D.E., PODACK, E.R., COOPER, N.R. (1980) J. Immunol., 124, 326.

KAPITANY, R.A., ZEBROWSKY, E.J. (1973) Anal. Biochem., 56, 361.

KASSEL, B., KAY, J. (1973) Science, 180, 1022.

KATAYAMA, K., ERICSSON, L.H., ENFIELD, D.L., WALSH, K.A., NEURATH, H., DAVIE, E.W., TITANI, K. (1979) Proc. Natl. Acad. Aci. U.S.A., 76, 4990.

KATZ, A.M., DREYER, W.J., ANFINSEN, C.B. (1959) J. Biol. Chem., 234, 2897.

KAY, J., KASSELL, B. (1971) J. Biol. Chem., 246, 6661.

KAZATCHKINE, M.D., FEARON, D.T., SILBERT, J.E., AUSTEN, K.F. (1979) J. Exp. Med., 150, 1202.

KEHOE, J.M., BOURGOIS, A., CAPRA, J.D., FOUGEREAU, M. (1974) Biochem., 13, 2499.

KEHOE, J.M., FOUGEREAU, M. (1969) Nature, 224, 1212.

KERR, M.A. (1979) Biochem. J., 183, 615.

KERR, M.A. (1980) Biochem. J., 189, 173.

KERR, M.A., PORTER, R.R. (1978) Biochem. J., 171, 99.

KERR, M.A., WALSH, K.A., NEURATH, H. (1975) Biochem., 14, 5088.

KLAPPER, D.G., WILDE, L-E, CAPRA, J.D. (1978) Anal. Biochem. 85, 126.

KNOBEL, H.R., HEUSSER, C., RODERICK, M.L., ISLIKER, H. (1974) J. Immunol., 112, 2094.

KNOBEL, H.R., VILLIGER, W., ISLIKER, H. (1975) Eur. J. Immunol., 5, 78.

KONNO, T., KATSUNO, Y., HIRAI, H. (1978) J. Immunol. Methods, 21, 325.

KRAUT, J. (1971) in "The Enzymes", 3, 165 (ed. P.D. Boyer) Acad. Press, New York.

KRAUT, J. (1977) Ann. Rev. Biochem., 46, 331.

KRIEGER, M., KAY, L.M., STROUD, R.M. (1974) J. Mol. Biol., 83, 209.

LACHMANN, P.J. (1973) in "Defence and Recognition", p.361 (ed. R.R. Porter) Butterworths, London.

LACHMANN, P.J. (1979) Behring Inst. Mitt., 63, 25.

LACHMANN, P.J., HALBWACHS, L. (1975) Clin. Exp. Immunol., 21, 109.

LACHMANN, P.J., HOBART, M.J. (1979) Ciba Foundation Symposium, 66, 231.

LACHMANN, P.J., MUNN, E.A., WEISSMANN, G. (1970) Immunol., 19, 983.

LAEMMLI, U.K. (1970) Nature, 227, 680.

LASKOWSKI, M., KATO, I. (1980) Ann. Rev. Biochem., 49, 593.

LAW, S.K., FEARON, D.T, LEVINE, R.P. (1979a) J. Immunol., 122, 759.

LAW, S.K., LEVINE, R.P. (1977) Proc. Natl. Acad. Sci. U.S.A., 74, 2701.

LAW, S.K., LICHTENBERG, N.A., LEVINE, R.P. (1979b) J. Immunol., 123, 1388.

LAWSON, W.B. (1980) Biochem., 19, 2140.

LAZDUNSKI, M., DELAAGE, M. (1965) Biochim. Biophys. Acta., 105, 541.

LEPOW, I.H., NAFF, G.B., PENSKY, J. (1965) Ciba Foundation Symposium, 57, 74 (ed. G.E.W. Wolstenholme and J. Knight) Churchill, London.

LESAVRE, P.H., GAILLARD, M.H., HALBWACHS, L. (1980) J. Immunol., 124, 1528 (abstract).

LESAVRE, P.H., HUGLI, T.E., ESSER, A.F., MÜLLER-EBERHARD, H.J. (1979) J. Immunol., 123, 529.

LESAVRE, P.H., MÜLLER-EBERHARD, H.J. (1978) J. Exp. Med., 148, 1498.

LIN, T-Y, FLETCHER, D.S. (1980) J. Biol. Chem., 255, 7756.

MAHONEY, W.C., HERMODSON, M.A. (1979) Biochem., 18, 3810.

MARTIN, A., LACHMANN, P.J., HALBWACHS, L., HOBART, M.J. (1976) Immunochem., 13, 317.

MATTHEWS, D.A., ALDEN, R.A., BIRKTOFT, J.J., FREER, S.T., KRAUT, J. (1977) J. Biol. Chem., 252, 8875.

MAYER, M.M. (1961) in "Experimental Immunochemistry" p.133 (ed. E.A. Kabat and M.M. Mayer). Thomas, Illinois.

MAYER, M.M. (1972) Proc. Natl. Acad. Sci. U.S.A., 69, 2954.

MAYER, M.M., HAMMER, C.H., MICHAELS, D.W., SHIN, M.L. (1979) Immunochem., 15, 813.

MEDICUS, R.G., GÖTZE, O., MÜLLER-EBERHARD, H.J. (1976) Scand. J. Immunol., 5, 1049.

METZGER, H. (1974) Adv. Immunol., 18, 169.

METZGER, H. (1978) Contemp. Topics in Mol. Immunol., 7, 119.

MICHAELS, D.W., ABRAMOVITZ, A.S., HAMMER, C.H., MAYER, M.M. (1976) Proc. Natl. Acad. Sci. U.S.A., 73, 2852.

MOON, A.Y., MERCUROFF, J., HESS, G.P. (1965a) J. Biol. Chem., 240, 717.

MOON, A.Y., STURTEVANT, J.M., HESS, G.P. (1965b) J. Biol. Chem., 240, 4204.

MOORE, G.W., GOODMAN, M. (1977) J. Mol. Evol., 9, 121.

MOORE, S., STEIN, W.H. (1963) Methods in Enzymology, 6, 819.

MÜLLER-EBERHARD, H.J. (1975) Ann. Rev. Biochem., 44, 697.

MÜLLER-EBERHARD, H.J., DALMASSO, A.P., CALCOTT, M.A. (1966) J. Exp. Med., 123, 33.

MÜLLER-EBERHARD, H.J., GÖTZE, O. (1972) J. Exp. Med., 135, 1003.

MÜLLER-EBERHARD, H.J., LEPOW, I.H. (1965) J. Exp. Med., 121, 819.

MÜLLER-EBERHARD, H.J., POLLEY, M.J., CALCOTT, M.A. (1967) J. Exp. Med., 125, 359.

NAGASAWA, S., STROUD, R.M. (1977) Immunochem., 14, 749.

NELSON, B., RUDDY, S. (1979) J. Immunol., 122, 1994.

NIEMANN, M.A., VOLANAKIS, J.E., MOLE, J.E. (1980) Biochem., 19, 1576.

NICOL, P.A.E., LACHMANN, P.J. (1973) Immunol., 24, 259.

NUTTAL, G. (1888) Z. Hyg., 4, 353.

OMENN, G.S., FONTANA, A., ANFINSEN, C.B. (1970) J. Biol. Chem., 245, 1895.

OPPENHEIMER, H.L., LABOUESSE, B., HESSE, G.P. (1966) J. Biol. Chem., 241, 2720.

PANGBURN, M.K., MORRISON, D.C., SCHREIBER, R.D., MULLER-EBERHARD, H.J. (1980) J. Immunol., 124, 977.

PÂQUES, E.P., HUBER, R., PRIESS, H. (1979) Hopper-Seyler's Z. Physiol. Chem., 360, 177.

PARKER, K.L., SHREFFLER, D.C., CAPRA, J.D. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 4275.

PENKE, B., FERENCZI, R., KOVAKS, K. (1974) Anal. Biochem., 60, 45.

PILLEMER, L. (1955) Trans. N.Y. Acad. Sci (2), 17, 526.

PILLEMER, L., BLUM, L., LEPOW, I.H., ROSS, O.A., TODD, E.W., WARDLAW, A.C. (1954) Science, 120, 279.

PILLEMER, L., BLUM, L., PENSKY, J., LEPOW, I.H. (1953) J. Immunol., 71, 331.

PODACK, E.R., BIESECKER, G., MÜLLER-EBERHARD, H.J. (1979) Proc. Natl. Acad. Sci. U.S.A., 76, 897.

PODACK, E.R., ESSER, A.F., BIESECKER, G., MÜLLER-EBERHARD, H.J. (1980) J. Exp. Med., 151, 301.

PODACK, E.R., MÜLLER-EBERHARD, H.J. (1978) J. Immunol., 121, 1025.

POLHILL, R.B., NEWMAN, S.L., PRUITT, K.M., JOHNSTON, R.B. (1978) J. Immunol., 121, 371.

POLLEY, M.J., MÜLLER-EBERHARD, H.J. (1967) J. Exp. Med., 126, 1013.

POLLEY, M.J., MÜLLER-EBERHARD, H.J. (1968) J. Exp. Med., 128, 533.

PORTER, R.R. (1979) in "Defence and Recognition, IIB. Structural Aspects." 23, 177 (ed. E.S. Lennox) University Park Press, Baltimore.

PORTER, R.R., REID, K.B.M. (1978) Nature, 275, 699.

PORTER, R.R., REID, K.B.M. (1979) Adv. Prot. Chem., 33, 1.

RADHAKRISHNAN, T.M., WALSH, K.A., NEURATH, H. (1967) J. Am. Chem. Soc., 89, 3059.

REID, K.B.M. (1976) Biochem. J., 155, 5.

REID, K.B.M. (1979) Biochem. J., 179, 367.

REID, K.B.M. (1980) in "Comprehensive Biochemistry" 19B, 461. (ed. M. Florken and E.H. Stotz) Elsevier, Amsterdam.

REID, K.B.M., LOWE, D.M., PORTER, R.R. (1972) Biochem. J., 130, 749.

REID, K.B.M., PORTER, R.R. (1976) Biochem. J., 155, 19.

REID, K.B.M., PORTER, R.R. (1981) Ann. Rev. Biochem., 50 (in press).

REID, K.B.M., SIM, R.B., FAIERS, A.P. (1977) Biochem. J., 161, 239.

REID, K.B.M., THOMPSON, E.O.P. (1978) Biochem. J., 173, 863.

RITZ, H. (1913) Z. Immunitätsforsch, 13, 62.

ROBSON, B. (1980) TIBS, 5, 240.

ROHRBACH, M.S., HUMPHRIES, B.A., YOST, F.J., RHODES, W.G., BOATMAN, S., HISKEY, R.G., HARRISON, J.H. (1973) Anal. Biochem., 52, 127.

ROSEN, F.S., ALPER, C.A. (1972) J. Clin. Invest., 51, 80a (abstract

SALVESON, G.S., BARRETT, A.J. (1980) Biochem. J., 187, 695.

SCHARFSTEIN, J., FERREIRA, A., GIGLI, I., NUSSENZWEIG, V. (1978) J. Exp. Med., 148, 207.

SCHECHTER, I., BERGER, A. (1967) Biochem. Biophys. Res. Commun., 27, 157.

SCHLEUNING, W-D, FRITZ, H. Methods in Enzymology, 45, 333.

SCHREIBER, R.D., GÖTZE, O., MÜLLER-EBERHARD, H.J. (1976) J. Exp. Med., 144, 1062.

SCHREIBER, R.D., PANGBURN, M.K., LESAVRE, P.H., MÜLLER-EBERHARD, H.J. (1978) Proc. Natl. Acad. Sci. U.S.A., 75, 3948.

SCHUMAKER, V.N., CALCOTT, M.A., SPIEGELBERG, H.L., MÜLLER-EBERHARD, H.J. (1976) Biochem. 15, 5175.

SEGAL, D.M., POWERS, J.C., COHEN, G.H., DAVIES, D.R., WILCOX, P.E. (1971) Biochem., 10, 3728.

SHELTON, E., YONEMASU, K., STROUD, R.M. (1972) Proc. Natl. Acad. Sci. U.S.A., 69, 65.

SHOTTON, D.M., WATSON, H.C. (1970) Nature, 225, 811.

SIM, R.B., ARLAUD, G.J., COLOMB, M.C. (1979) Biochem. J. 179, 449.

SIM, R.B., PORTER, R.R., REID, K.B.M., GIGLI, I. (1977) Biochem. J., 163, 219.

SIM, R.B., TWOSE, T.M., SIM, E., PATERSON, D.S., Biochem. J. (in press).

SISSONS, J.G.P., SCHREIBER, R.D., PERRIN, L.H., COOPER, N.R., MÜLLER-EBERHARD, H.J., OLDSTONE, M.B.A. (1979) J. Exp. Med., 150, 445.

SOBER, H.A. (1970) "Handbook of Biochemistry" p.C42 (Chemical Rubber Co., Cleveland, Ohio).

SPIEGELBERG, H.L. (1974) Adv. Immunol., 19, 259.

STEITZ, T.A., HENDERSON, R., BLOW, D.M. (1969) J. Mol. Biol., 46, 337.

STOLFI, R.L. (1968) J. Immunol., 100, 46.

STROUD, R.M., KAY, L.M., DICKERSON, R.E. (1974) J. Mol. Biol., 83, 185.

STROUD, R.M., KRIEGER, M., KOEPPE, R.E., KOSSIAKOFF, A.A., CHAMBERS, J.L. (1975) in "Proteases and Biological Control" p.13 (ed. E. Reich, D.B. Rifkin and E. Shaw). Cold Spring Harbor Laboratory.

STRUNK, R.C., GICLAS, P.C. (1980) J. Immunol., 124, 520.

SUNDSMO, J.S., CURD, J.G., KOLB, W.P., MÜLLER-EBERHARD, H.J. (1978b) J. Immunol., 120, 855.

SUNDSMO, J.S., GÖTZE, O. (1980) Cell Immunol., 52, 1.

SUNDSMO, J.S., KOLB, W.P., MÜLLER-EBERHARD, H.J. (1978a) J. Immunol., 120, 850.

TAKAHASHI, M., CZOP, J., FERREIRA, A., NUSSENZWEIG, V. (1976) Transpl. Rev., 32, 121.

TAMURA, H., NELSON, R.A. (1967) J. Immunol., 99, 582.

THOMPSON, R.C., BLOUT, E.R. (1970) Proc. Natl. Acad. Sci. U.S.A., 67, 1734.

VILLIERS, C.L., ARLAUD, G.A., PAINTER, R.H., COLOMB, M.G. (1980) FEBS Letters, 117, 289.

VOGT, W., DAMES, W., SCHMIDT, G., DIEMINGER, L. (1977) Immunochem., 14, 201.

VOGT, W., DIEMINGER, L., LYNEN, R., SCHMIDT, G. (1974) Hoppe-Seyler's Z. Physiol. Chem., 355, 171.

VOGT, W., SCHMIDT, G., von BUTTLAR, B., DIEMINGER, L. (1978) Immunol., 34, 29.

VOGT, W., SCHMIDT, G., DIEMINGER, L., LYNEN, R. (1975) Z. Immun-Forsch., 149, 440.

VOGT, W., SCHMIDT, G., HINSEL, B. (1979) Immunol., 36, 139.

VOLANAKIS, J.E., BHOWN, A.S., BENNETT, J.C., MOLE, J.E. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 1116.

VOLANAKIS, J.E., SCHROHENLOHER, R.E., STROUD, R.M. (1977) J. Immunol., 119, 337.

WALSH, K.A. (1970) Methods in Enzymology, 19, 41.

WATSON, H.C., SHOTTON, D.M., COX, J.M., MUIRHEAD, H. (1970) Nature, 225, 806.

WAUTIER, J-L, SOUCHON, H., REID, K.B.M., PELTIER, A.P., CAEN, J.P. (1977) Immunochem., 14, 763.

WEBER, K., OSBORN, M. (1969) J. Biol. Chem., 244, 4406.

WEBER, K., OSBORN, M. (1975) in "The Proteins" 1, 179 (ed. N. Neurath and R.L. Hill) Acad. Press, New York.

WEILER, J.M., DAHA, M.R., AUSTEN, K.F., FEARON, D.T. (1976) Proc. Natl. Acad. Sci. U.S.A., 73, 3268.

WHALEY, K. (1980) J. Exp. Med., 151, 501.

WHALEY, K., RUDDY, S. (1976) J. Exp. Med., 144, 1147.

WHALEY, K., THOMPSON, R.A. (1978) Immunol., 35, 1045.

WHITEHEAD, R.H., GORDON, J., WORMALL, A. (1925) Biochem. J., 19, 618.

WINDHAGER, E.E. (1969) Ann. Rev. Physiol., 31, 117.

WOODBURY, R.G., KATUNUMA, N., KOBAYASHI, K., TITANI, K., NEURATH, H. (1978) Biochem., 17, 811.

WOODS, D.R., WANG, K-T, (1967) Biochim. Biophys. Acta., 133, 369.

WRIGHT, H.T. (1973) J. Mol. Biol., 79, 13.

YASMEEN, D., ELLERSON, J.R., DORRINGTON, K.J., PAINTER, R.H. (1976) J. Immunol., 116, 518.

von ZABERN, I., MOLTE, R., VOGT, W. (1980) Scand. J. Immunol., 9, 69.

ZICCARDI, R.J., COOPER, N.R. (1976) J. Immunol., 116, 498.

ZICCARDI, R.J., COOPER, N.R. (1979) J. Immunol., 123, 788.

Biochem. J. (1980) **187**, 863–874
Printed in Great Britain

Factor D̄ of the Alternative Pathway of Human Complement

PURIFICATION, ALIGNMENT AND *N*-TERMINAL AMINO ACID SEQUENCES OF THE MAJOR CYANOGEN BROMIDE FRAGMENTS, AND LOCALIZATION OF THE SERINE RESIDUE AT THE ACTIVE SITE

Diana M. A. JOHNSON, Jean GAGNON and Kenneth B. M. REID
Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(*Received 20 November 1979*)

1. The serine esterase factor D̄ of the complement system was purified from outdated human plasma with a yield of 20% of the initial haemolytic activity found in serum. This represented an approx. 60000-fold purification. 2. The final product was homogeneous as judged by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (with an apparent mol.wt. of 24000), its migration as a single component in a variety of fractionation procedures based on size and charge, and its *N*-terminal amino-acid-sequence analysis. 3. The *N*-terminal amino acid sequence of the first 36 residues of the intact molecule was found to be homologous with the *N*-terminal amino acid sequences of the catalytic chains of other serine esterases. Factor D̄ showed an especially strong homology (>60% identity) with rat 'group-specific protease' [Woodbury, Katunuma, Kobayashi, Titani, & Neurath (1978) *Biochemistry* **17**, 811–819] over the first 16 amino acid residues. This similarity is of interest since it is considered that both enzymes may be synthesized in their active, rather than zymogen, forms. 4. The three major CNBr fragments of factor D̄, which had apparent mol.wts. of 15800, 6600 and 1700, were purified and then aligned by *N*-terminal amino acid sequence analysis and amino acid analysis. By using factor D̄ labelled with di-[1,3-^{14}C]isopropylphosphofluoridate it was shown that the CNBr fragment of apparent mol.wt. 6600, which is located in the C-terminal region of factor D̄, contained the active serine residue. The amino acid sequence around this residue was determined.

Factor D̄ is an enzyme of the alternative pathway of complement that has some of the characteristics of a typical serine esterase. It is composed of a catalytic chain of approx. 23000–25000 mol.wt. (Götze, 1976; Volanakis *et al.*, 1977) and is irreversibly inhibited by iPr_2P-F (Fearon *et al.*, 1974). It has an *N*-terminal amino acid sequence of Ile-Leu-Gly-Gly- (Volanakis *et al.*, 1977; Davis *et al.*, 1979*a*,*b*) which is very characteristic of the type of highly conserved amino-acid sequences found at the *N*-terminal end of the catalytic chains of other serine esterases (de Haën *et al.*, 1975). It is not known whether factor D̄ is synthesized first as a zymogen or in its active form. There is one report that a small proportion of the enzyme present in the plasma is in the proenzyme form and that conversion of this material to the active enzyme can be achieved by trypsin treatment (Fearon *et al.*, 1974). However recent studies suggest that it is probable that factor D̄ may normally be present in plasma only in its activated form (Lesavre & Müller-Eberhard, 1978). Factor B is another serine-esterase component of the alternative pathway that has clearly been shown to be present in plasma in its zymogen form. When factor B is complexed with C3b, in the presence of Mg^{2+}, it is readily split by factor D̄ at a single Arg–Lys bond thus yielding the Ba and Bb fragments (Lesavre *et al.*, 1979). This limited proteolysis of Factor B brings about the formation of the active enzyme complex $\overline{C3b\ Bb}$ which can split and activate C3, the most abundant complement component, which plays a central role in the activation of the complement system (Schreiber *et al.*, 1978; Austen, 1978). The initial activation steps of the alternative pathway are not

Abbreviations used: SDS, sodium dodecyl sulphate; iPr_2P-F, di-isopropyl phosphorofluoridate; Quadrol, *NNN'N'*-tetrakis-(2-hydroxypropyl)ethenediamine trifluoroacetate; Polybrene, 1,5-diaza-1,5-dimethylundecamethylene polymethobromide.

0306-3275/80/060863-12$01.50/1

fully understood and structural studies may help clarify the role that factor D̄ plays in these steps.

The study of factor D̄ has been hampered by its low serum concentration (approx. 1–2 mg/litre) and by the difficulty in removing contaminants of similar apparent molecular weight and behaviour on ion-exchange columns. This paper describes the isolation of highly purified factor D̄ in sufficient quantities to carry out *N*-terminal sequence analysis on the intact molecule, and purification and *N*-terminal sequence analysis of the peptides generated by CNBr treatment of the reduced and alkylated intact molecule.

The results obtained confirm, and clarify, the *N*-terminal sequence obtained for the first nine residues reported by Davis *et al.* (1979*a*,*b*) and extend the *N*-terminal sequence information to residue 36. Amino-acid analysis and amino-acid sequencing has also allowed the alignment of the three major peptides produced by CNBr treatment. Labelling of factor D̄ with [1,3-^{14}C]iPr_2P-F has allowed the identification and localization of the serine residue at the active site.

Materials and Methods

Materials

Fresh frozen human plasma and outdated human plasma, which had been collected in sodium citrate anticoagulant, was obtained from the Churchill Hospital, Oxford. Plasma was clotted overnight, at 4°C, by the addition of 1.0 M-$CaCl_2$ to give a final concentration of 20 mM. The clot was removed by centrifugation and filtration through muslin, and the serum was stored at −20°C.

Fresh frozen plasma was clotted and used for the preparation of factor D̄-deficient serum as described by Martin *et al.* (1976), using a Sephadex G-75 column equilibrated with 0.1 M-Tris/HCl/0.2 M-NaCl/2 mM-EDTA, pH 8.0 at 4°C.

CM-Sephadex C-50, Sephadex G-75, concanavalin A–Sepharose, Sephadex G-50 (superfine) and Sephadex G-75 (superfine) were all obtained from Pharmacia Fine Chemicals, Uppsala, Sweden. CM-cellulose 32 was obtained from Whatman, Maidstone, Kent, U.K.

Ovalbumin, bovine α-chymotrypsinogen A, whale skeletal muscle myoglobin, egg white lysozyme, horse-heart cytochrome *c* and bovine insulin were all obtained from Sigma. Bovine serum albumin was obtained from Armour Pharmaceutical Co., Eastbourne, Sussex, U.K. Iodoacetamide, iodoacetic acid and dithiothreitol were obtained from BDH. CNBr was from Koch–Light Laboratories, Colnbrook, Bucks., U.K.

Quadrol was purchased from Beckman Instruments, Palo Alto, CA, U.S.A. Polybrene was from Aldrich Chemical Co., R.N. Emanuel, Wembley, Middlesex, U.K. All other chemicals used in the sequencer were obtained from Rathburn Chemicals, Walkerburn, Peebleshire, Scotland, U.K.

Iodo[2-^{14}C]acetic acid (54 Ci/mol) was from The Radiochemical Centre, Amersham, Bucks., U.K. [1,3-^{14}C]iPr_2P-F (100 Ci/mol) was purchased from New England Nuclear.

Haemolytic assay for factor D̄

Factor D̄ was assayed using the haemolytic diffusion plate assay of Martin *et al.* (1976). The assay was standardized by using dilutions of pooled normal serum. One unit of factor D̄ activity was defined as the haemolytic area obtained using 10 μl of the standard serum.

Purification of factor D̄

Human serum (2000 ml) was dialysed for 20 h at 4°C against 20 litres of 5 mM-EDTA, pH 5.4. The euglobulin precipitate was removed by centrifugation at 2000 *g* for 1 h. The conductivity of the pseudoglobulin solution was adjusted to 12–14 mS using 1 M-sodium phosphate buffer, pH 6.0, and the solution was loaded on to a column (12 cm × 8 cm) of CM-Sephadex C-50, equilibrated in 0.2 M-sodium phosphate buffer, pH 6.0 (6.15 ml of 0.2 M-Na_2HPO_4 + 43.85 ml of 0.2 M-NaH_2PO_4/50 ml of buffer). The column was washed with 1500 ml of 0.2 M-sodium phosphate, pH 6.0, and then the factor D̄ was eluted using 2000 ml of 0.4 M-sodium phosphate buffer, pH 6.0 (12.3 ml of 0.4 M-Na_2HPO_4 + 87.7 ml of 0.4 M-NaH_2PO_4/100 ml of buffer. The flow rate was 600 ml/h during application of the 0.2 M-sodium phosphate wash, and 50 ml/h during the elution of factor D̄.

The fractions containing factor D̄ haemolytic activity were pooled and $(NH_4)_2SO_4$ was added to 50% saturation (291 g/litre). The suspension was stirred for 2 h at 4°C, then centrifuged at 10000 *g* for 2 h. $(NH_4)_2SO_4$ was added to the supernatant (125 g/litre) to give 70% saturation and the suspension was stirred for 2 h at room temperature. The precipitate was dissolved in 40 ml of 0.1 M-Tris/HCl/0.2 M-NaCl/2 mM-EDTA, pH 8.0 at 4°C, and, after centrifugation, applied to a column (5 cm × 100 cm) of Sephadex G-75 equilibrated with the same buffer. The column was run at a flow rate of 30 ml/h. The fractions containing factor D̄ haemolytic activity were pooled and concentrated to 10 ml using an Amicon Diaflo Ultrafiltration cell fitted with a UM2 membrane.

These initial steps in the purification procedure were developed by R. Prohaska and provide a factor D̄ preparation of sufficient purity for functional studies. However, for structural studies, further purification of the factor D̄ was considered necessary.

The material from the Sephadex G-75 steps from four different preparations was pooled and dialysed against 0.23 M-acetic acid/NaOH buffer, pH 5.2 (5.5 ml of 4 M-NaOH adjusted to pH 5.2 with acetic acid and made up to 100 ml) and applied to a column (1.5 cm × 20 cm) of CM-cellulose 32 that had been equilibrated with the same buffer. The column was washed with the starting buffer at 30 ml/h until the A_{230} of the eluate was zero, and then a linear gradient was developed using 120 ml of 0.23 M-acetic acid/NaOH buffer, pH 5.2, and 120 ml of the same buffer containing 0.3 M-NaCl. The fractions containing factor D̄ haemolytic activity were pooled and concentrated to 10 ml, then dialysed against 0.1 M-sodium acetate/acetic acid buffer (0.1 M-sodium acetate adjusted to pH 6.0 with acetic acid) containing 1 M-NaCl, 1 mM-$MgCl_2$, 1 mM-$MnCl_2$ and 1 mM-$CaCl_2$. After dialysis the sample (10 ml) was applied to a column (1 cm × 12 cm) of concanavalin A–Sepharose, equilibrated with the same buffer, and eluted at a flow rate of 10 ml/h. The factor D̄ haemolytic activity was not retarded on the column. A contaminant, of the same apparent molecular weight as factor D̄ in non-reducing conditions on SDS/polyacrylamide gels, was bound to the column and could be eluted with 2.5% (w/v) 1-*O*-methyl-α-D-glucopyranoside dissolved in the same buffer used to equilibrate the column.

The factor D̄ obtained from the concanavalin A–Sepharose column was dialysed extensively against 0.5 M-acetic acid and then was freeze-dried prior to use in the sequencing studies.

SDS/polyacrylamide gel electrophoresis and determination of the apparent molecular weights of factor D̄ and its CNBr fragments

Electrophoresis was carried out as described by Laemmli (1970) using 15% or 20% (w/v) polyacrylamide slab gels containing 1% (w/v) SDS. Proteins were stained using Coomassie Brilliant Blue.

Non-reduced protein samples (5–10 μg) in 20 mM-iodoacetamide/0.1 M-Tris/HCl/4 M-urea/1% (w/v) SDS, pH 8.0, were incubated for 2 min at 100°C. Reduced samples were prepared by incubation with 20 mM-dithiothreitol for 2 min at 100°C.

The apparent molecular weights of factor D̄ and its CNBr fragments were calculated by comparison with the mobilities of reduced and non-reduced marker proteins and peptides: bovine serum albumin (mol.wt. 67000), ovalbumin (mol.wt. 42000), bovine pancreas α-chymotrypsinogen A (mol.wt. 25700), whale skeletal muscle myoglobin (mol.wt. 17200), egg-white lysozyme (mol.wt. 14300), horse-heart cytochrome *c* (mol.wt. 12400), bovine insulin (mol.wt. 5800), B chain of insulin (mol.wt. 3400) and A chain of insulin (mol.wt. 2500).

Reduction and alkylation of factor D̄

Factor D̄ (90 nmol) was dissolved in 6 M-guanidine/HCl/0.4 M-Tris/HCl/2 mM-EDTA, pH 8.0 (1.5 ml) containing 20 mM-dithiothreitol, and incubated in a sealed flask at 37°C for 3 h with frequent mixing. Iodo[2-^{14}C]acetic acid (100 μCi in 100 μl of distilled water) was added and the incubation was continued for 5 min at 25°C; then ice-cold carrier iodoacetic acid was added to a concentration of 60 mM, and the mixture was cooled on ice for 45 min. The reduced and alkylated sample was then dialysed at 4°C against distilled water (1 litre) for 3 h, followed by two changes of 0.5 M-acetic acid (1 litre) and was freeze-dried.

Labelling of factor D̄ with [1,3-^{14}C]iPr_2P-F

Factor D̄ (50 nmol) in 5 ml of 0.01 M-Tris/HCl/0.15 M-NaCl/1 mM-$MgCl_2$/1 mM-$CaCl_2$, pH 8.0 at 4°C, was incubated at 30°C for 1 h with 1 μmol of [1,3-^{14}C]iPr_2P-F (100 μCi). Non-radioactive iPr_2P-F was added to give a final concentration of 10 mM and the mixture was incubated for a further 1 h at 30°C. The inactivated factor D̄ was dialysed for 24 h against three changes of 0.5 M-acetic acid and was freeze-dried. Radioactive samples [in 10–30 μl of 5% (v/v) formic acid] were added to 10 ml of 1,4-dioxan containing 2% (w/v) naphthalene and 0.5% (w/v) 2,5-diphenyloxazole. Samples were counted for radioactivity using an LKB-Wallac 1210 Ultrobeta counter.

CNBr digestion of factor D̄ and purification of the major CNBr peptides

Reduced and alkylated factor D̄ (50 nmol) was dissolved in 70% (v/v) formic acid (0.7 ml) containing CNBr (0.15 mmol) and incubated for 16 h at 25°C. The mixture was freeze-dried and then redissolved in 70% (v/v) formic acid (0.7 ml) and applied to a column (1.5 cm × 100 cm) of Sephadex G-50 (superfine) equilibrated with 5% (v/v) formic acid. Peptides eluting near the void volume of the Sephadex G-50 column were further purified by gel filtration on a column (1.5 cm × 100 cm) of Sephadex G-75 (superfine) equilibrated with 10% (v/v) formic acid. Smaller peptides were further purified by high pressure liquid chromatography using a μBondapak C-18 column. Two solutions were used in the high pressure liquid chromatography of peptides: solution A, which was composed of CH_3CN and 10 mM-NH_4HCO_3 (5:95, v/v) and solution B, which was composed of CH_3CN and 10 mM-NH_4HCO_3 (64:40, v/v). The column was equilibrated with a mixture of solution A and solution B in the ratio 95:5 (v/v) and then eluted with a linear gradient to give a final ratio of 10:90 (v/v) of solution A to solution B. The column was pumped at 1.5 ml/min by a Waters Associates system (model 6000 A).

Amino acid analysis

Proteins and peptides were hydrolysed, under vacuum, at 110°C for 48h in constant-boiling HCl containing 4mM-phenol. No corrections were made for destruction of threonine and serine, and tryptophan was not measured. Cystine was measured as *S*-carboxymethylcysteine after acid hydrolysis, under vacuum, of reduced and alkylated samples in the presence of 0.05% (v/v) 2-mercaptoethanol.

Automated amino-acid sequence determination

The *N*-terminal amino acid sequences of factor D̄ and its CNBr fragments were determined by automated Edman degradation in a Beckman 890c sequencer using the 0.3M-Quadrol programme of Hunkapiller & Hood (1978). Polybrene (2mg) was added to the sequencer cup prior to application of each sample (Klapper *et al.*, 1978). The thiazolinones released were converted into the amino-acid phenylthiohydantoin derivatives by heating at 80°C under N_2 for 10min in 1M-HCl (200μl) containing 0.1% (v/v) ethanethiol. After two extractions with ethyl acetate (700μl), 40–80% of the extracted amino acid phenylthiohydantoin derivative was examined by high-pressure liquid chromatography. The chromatography was done on a μBondapak C-18 column using a 14–56% methanol gradient in aqueous buffer (Bridgen *et al.*, 1976) pumped at 2.2ml/min by a Waters Associates system (model 6000A) with a u.v. detector.

Results

Purification of factor D̄

A summary of the purification of factor D̄ is given in Table 1. After precipitation of euglobulin, more than 99% of the protein remaining in the pseudoglobulin fraction was removed by ion-exchange chromatography on CM-Sephadex C-50. On washing the column with 0.4M-sodium phosphate buffer, pH6.0, the factor D̄ haemolytic activity eluted above 16mS. The activity peak was broad and no improvement in purification resulted from eluting factor D̄ with a linear gradient instead of the stepwise wash. It was found convenient to concentrate the material from the CM-Sephadex C-50 column by $(NH_4)_2SO_4$ fractionation. Most of the factor D̄ haemolytic activity precipitated in the 50–70% saturation fraction.

The $(NH_4)_2SO_4$ precipitate was further purified by gel filtration on Sephadex G-75, where the factor D̄ haemolytic activity was separated from a large number of proteins all having apparent molecular weights >30000. Because of the small amount of protein (usually 7–10mg) in the factor D̄ pool from the Sephadex G-75 column it was found convenient to combine the Sephadex G-75 factor D̄ pools from four preparations before carrying out the final stages of the purification. The partially purified factor D̄ from Sephadex G-75 was therefore stored at −70°C until use.

Following gel filtration on Sephadex G-75, the preparation appeared to be functionally pure with respect to factor D̄ haemolytic activity, but on further purification on CM-cellulose 32 at least four protein peaks were eluted (Fig. 1*a* and *b*). The factor D̄ haemolytic activity eluted between 22–27mS. Another protein, which was removed by use of concanavalin-A–Sepharose, was eluted between 25–28mS in the gradient on CM-cellulose 32 (Fig. 1). This protein could not be distinguished from factor D̄ in non-reducing conditions on SDS/polyacrylamide gels, but in reducing conditions had a higher apparent molecular weight compared to factor D̄ (Fig. 1*b*). No factor D̄ haemolytic activity was associated with this contaminating protein, the concentration of which varied considerably from preparation to preparation, sometimes comprising approx. 50% of the protein in the partially purified factor D̄ pool from CM-cellulose 32. Thus the CM-cellulose 32 pool was routinely applied to a column of concanavalin-A–Sepharose. Factor D̄

Table 1. *Purification of factor D̄ from serum*

Fraction	Volume (ml)	Total protein (A_{280})	10^{-3} × total haemolytic activity (units)	Specific activity (units/A_{280})	Yield of activity (%)	Purification factor
Serum	2000	110600	191	1.73	100	1.0
Pseudoglobulin	2100	104300	176	1.69	92	0.98
CM-Sephadex C-50	900	211	160	758	84	440
50–70% $(NH_4)_2SO_4$ precipitate	—	72.4	130	1790	68	1040
Sephadex G-75	350	8.52	86.5	10200	45	5890
CM-cellulose-32	86	1.56	46.2	29600	24	17100
Concanavalin-A–Sepharose	24	0.875*	43.1	49200	22	28500

* The final yield of protein as estimated by amino-acid analysis was 0.42mg. This gives a final specific activity (units/mg of protein) for the factor D̄ of 102600 and a purification factor of approx. 60000.

was not retained on this column, but the contaminant was bound and could only be eluted with 1-*O*-methyl-α-D-glucopyranoside.

The purified factor D̄ appeared to be homogeneous on SDS/polyacryamide gels in reducing and non-reducing conditions (Fig. 2). On staining the gels with Coomassie Blue, the protein band appeared broad in the non-reduced sample, but amino-acid sequence analysis has shown the presence of only one sequence and the number of fragments obtained after CNBr treatment is also consistent with there being only one type of polypeptide chain in the final preparation.

Stability of factor D̄

Factor D̄ in 0.1 M-Tris/HCl/0.2 M-NaCl/2 mM-EDTA, pH 8.0, containing 0.02% (w/v) NaN_3, was stable at 4°C for several months, with only a slight decrease in haemolytic activity. However, on storage at this temperature, protein aggregation has been observed.

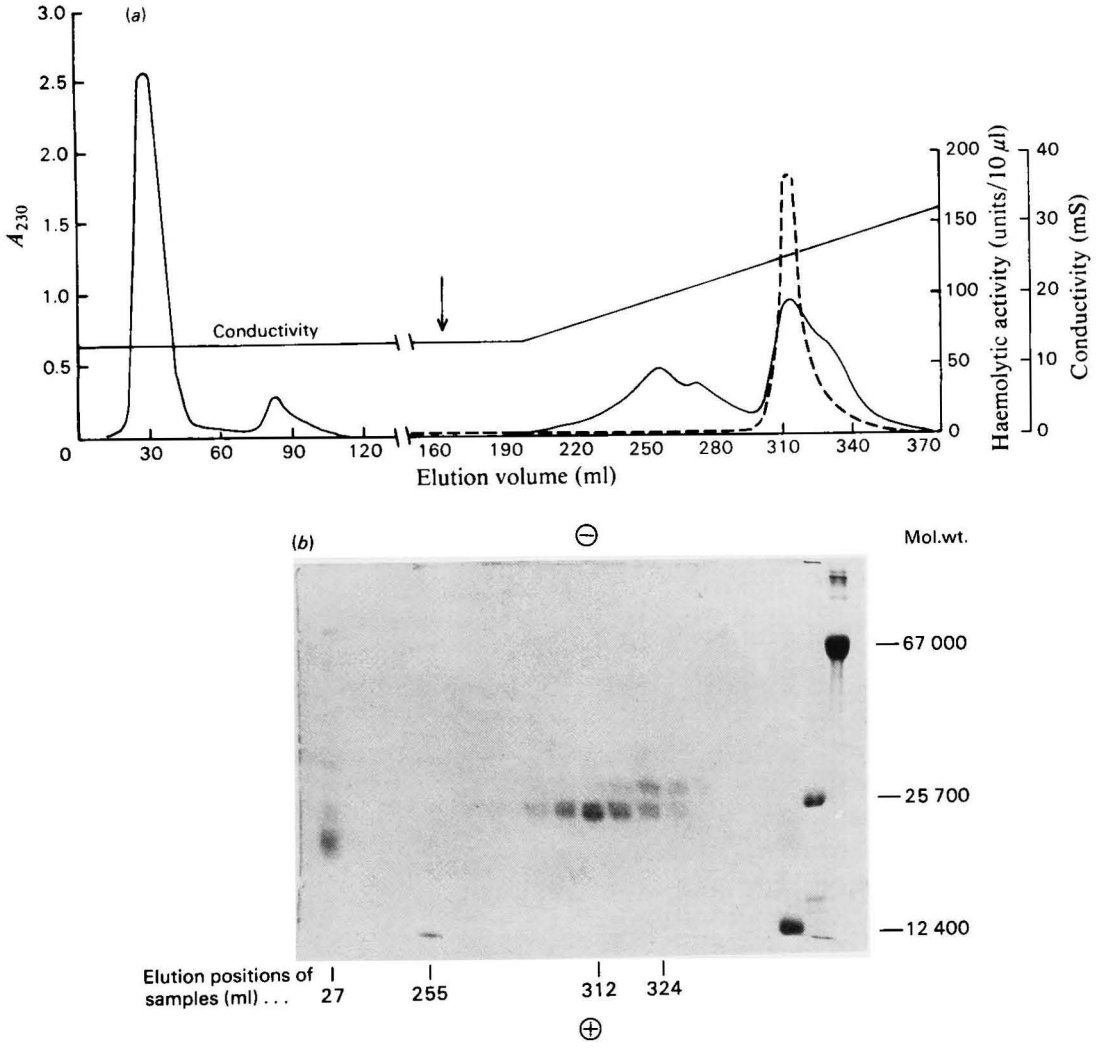


Fig. 1. *Ion-exchange chromatography of factor D̄ on CM-cellulose 32, and SDS/polyacrylamide-gel electrophoresis (Laemmli, 1970) of the fractions*

(*a*) The partially purified factor D̄ (approx. 30 mg) from Sephadex G-75 was equilibrated with 0.23 M-acetic acid/NaOH buffer, pH 5.2, and loaded on to a column (1.5 cm × 20 cm) of CM-cellulose 32 equilibrated in the same buffer. The column was developed as described in the Materials and Methods section. Factor D̄ haemolytic activity eluted between 22–27 mS. ——, A_{230}; – – –, haemolytic activity. The arrow marks the start of the gradient. (*b*) Samples (25 μl) were taken from column fractions corresponding to elution at 27, 81, 210, 240, 255, 270, 285, 294, 300, 306, 312, 318, 324, 330, 345 and 360 ml on the chromatogram profile shown in (*a*). The samples were reduced and alkylated as described in the Methods and run on a SDS/15% (w/v) polyacrylamide slab gel. The major contaminants in the preparation were eluted at peak volumes of 27, 255 and 324 ml and factor D was eluted at 312 ml. Albumin (mol.wt. 67 000), α-chymotrypsinogen A (mol.wt. 25 700) and myoglobin (mol.wt. 17 200) were run as marker proteins, after reduction and alkylation.

Storage at −70°C prevents aggregation, but freezing and thawing was found to cause a loss of approx. 30% of the original haemolytic activity.

Molecular weight determinations

The apparent molecular weight of factor D̄ was determined on 15% (w/v) polyacrylamide gels containing 1% (w/v) SDS. In non-reducing conditions the molecular weight was determined as 24000 and in reducing conditions as 23700. Reduced and alkylated factor D̄ had a slightly slower mobility on SDS/polyacrylamide gels, compared with non-reduced factor D̄ (Fig. 2).

CNBr digestion of factor D̄

From the methionine content of factor D̄ (1.2 methionine residues/100 residues) three, or possibly four, peptides were expected. The CNBr digest was initially fractionated on Sephadex G-50 (superfine) (Fig. 3). Two major peaks of protein, containing radioactivity, eluted at 56–76 ml and at 90–104 ml respectively. The second peak (90–104 ml) contained peptide CNBr-2. The first peak (56–76 ml) was found to be heterogeneous on SDS/20% (w/v) polyacrylamide gels and was repurified on a column (1.5 cm × 100 cm) of Sephadex G-75 (superfine) equilibrated in 10% (v/v) aq. formic acid. Peptide CNBr-1 eluted between 53–63 ml on this column. Peptide CNBr-3 was purified from the pool made between 105–150 ml on the initial fractionation of the digest on the Sephadex G-50 (superfine) column (Fig. 3). This pool was freeze-dried and then redissolved in 10 mM-NH_4HCO_3/5% (w/v) CH_3CN (50 μl) plus 99% formic acid (10 μl). The material in the pool was then further purified by high pressure liquid chromatography using a μBondapak C-18 column. Two significant peaks of peptide material were detected in an A_{210} ratio of approx. 10:1. The amino-acid compositions of these peptides were identical, suggesting that they were two forms of the same peptide. It is possible that one of the peptides had its *C*-terminal homoserine in the lactone form.

The apparent molecular weights of peptides CNBr-1 and CNBr-2, as estimated on SDS/20% (w/v) polyacrylamide gels, were 15800 and 6600. Peptide CNBr-3 was not detected on SDS/20% (w/v) polyacrylamide gels, but its molecular weight was estimated to be 1700 from its amino-acid composition (Table 2) and its amino-acid sequence (Fig. 4). The combined molecular weights of these three major CNBr peptides thus account for the total apparent molecular weight of factor D̄, although the possible occurrence of a fourth, very small, CNBr peptide has not been excluded.

Two other fragments were detected in CNBr digests of factor D̄. One of these (of apparent mol.wt. 18000) eluted on Sephadex G-75 superfine just ahead of peptide CNBr-1, from which it was not clearly resolved. This peptide, which was present at about one-fifth of the concentration of peptide CNBr-1, probably resulted from incomplete digestion of the methionyl bond at position 15 in factor D̄

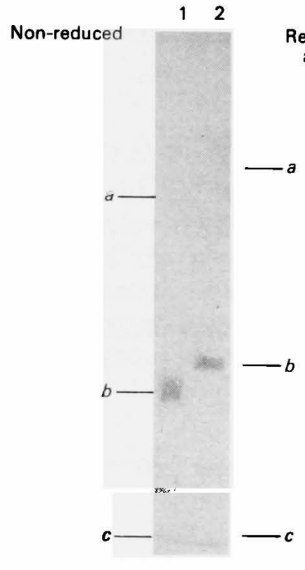


Fig. 2. *SDS/polyacrylamide-gel electrophoresis (Laemmli, 1970) of purified factor D*

Factor D̄ (5–10 μg), after the final purification step on concanavalin-A–Sepharose, was reduced and alkylated as described in the Materials and Methods section. Non-reduced (track 1) and reduced (track 2) samples were run on a 15% (w/v) SDS/polyacrylamide slab gel. The arrows show the positions of the following marker proteins: (*a*) albumin (mol.wt. 67000), (*b*) -chymotrypsinogen A (mol.wt. 25700), (*c*) cytochrome *c* (mol.wt. 12400), before and after reduction and alkylation.

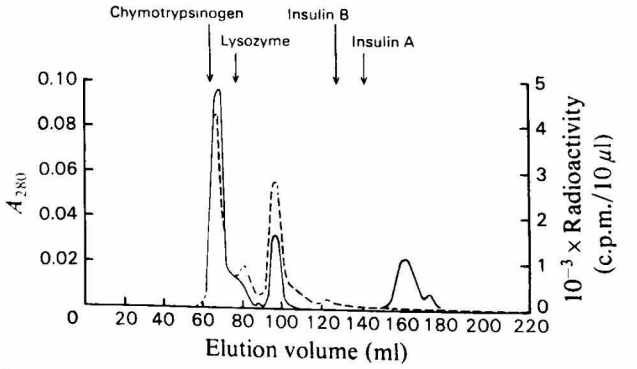


Fig. 3. *Purification on Sephadex G-50 of peptides obtained by CNBr digestion of factor D*

The CNBr digest was freeze-dried, redissolved in 70% (v/v) formic acid (0.7 ml) and applied to a column (1.5 cm × 100 cm) of Sephadex G-50 (superfine) equilibrated with 5% (v/v) formic acid and run at a flow rate of 2.5 ml/h. ——, A_{280}; – – –, radioactivity of *S*-[^{14}C]carboxymethylcysteine.

(Fig. 4) and would be composed of peptides CNBr-3 and CNBr-1. The other fragment, also present in low amounts, had an apparent mol.wt. of 23000. This fragment was distinguishable from residual uncleaved factor D̄ and is probably composed of peptides CNBr-1 and CNBr-2. Amino-acid sequence analysis on these two fragments was not performed.

Amino acid compositions, and N-terminal amino acid sequences, of factor D̄ and its CNBr fragments

The amino-acid compositions of factor D̄, peptide CNBr-1, peptide CNBr-2 and peptide CNBr-3 are given in Table 2.

As expected from earlier work (Volanakis *et al.*, 1977) the *N*-terminal amino acid of factor D̄ was found to be isoleucine. The *N*-terminal amino-acid sequences of factor D̄ and peptide CNBr-3 were found to be identical for the first 11 residues, except for residue 10, which was not identified in the sequencer run of peptide CNBr-3 (Table 3). The amino acid at residue 10 was shown to be histidine in the sequencer run of intact factor D̄ and this was consistent with the presence of one histidine residue in the amino-acid composition of peptide CNBr-3 (Table 2).

Table 2. *Amino-acid compositions of factor D̄ and its major CNBr peptides*

The amino acid compositions are given as residues/100 residues except in the case of peptide CNBr-3 which is given as mol of residue/mol of peptide. The samples were hydrolysed at 110°C for 47h. No corrections were made for serine and threonine destruction, and tryptophan was not determined. Cystine was estimated as *S*-carboxymethylcysteine.

Amino acid	Amino acid composition of: Factor D̄ (residues/100 residues)	CNBr-1 (residues/100 residues)	CNBr-2 (residues/100 residues)	CNBr-3 (mol/mol)
Asp	9.4	10.6	10.0	0.3 (0)
Thr	3.5	3.7	3.5	0.3 (0)
Ser	5.0	5.4	7.6	—
Hse	—	0.6	—	0.7 (1)
Glu	9.4	9.8	7.9	2.1 (2)
Pro	8.1	7.3	6.3	1.1 (1)
Gly	10.4	10.2	14.3	2.2 (2)
Ala	8.8	10.6	9.0	2.7 (3)
Val	8.0	8.3	9.8	0.3 (0)
Cys	4.0	1.6	3.0	—
Met	1.2	—	—	—
Ile	2.6	2.2	4.3	0.8 (1)
Leu	10.3	14.1	6.8	1.0 (1)
Tyr	1.8	0.4	2.4	0.9 (1)
Phe	1.3	0.2	0.7	—
His	5.2	6.5	1.9	1.0 (1)
Lys	3.9	2.1	4.9	—
Arg	7.1	6.4	7.6	1.6 (2)

From the apparent mol.wt. (15800) of peptide CNBr-1 it was estimated that the peptide was approx. 141 amino acid residues long. The *N*-terminal amino acid of peptide CNBr-1 was found to be alanine. By comparing the *N*-terminal amino acid sequence of peptide CNBr-1 with that of intact factor D̄ (Table 3) it can be seen that the *N*-terminal amino acid sequence of peptide CNBr-1 overlaps with that of factor D̄ from residue 16 onward. The only exception was at residue 11 of peptide CNBr-1, where *S*-carboxymethylcysteine was found compared with an unidentified residue in factor D̄ (Table 3). However this would be expected since the intact factor D̄ used for amino acid sequence analysis was not reduced and alkylated.

From the sequence data obtained for peptides CNBr-1 and CNBr-3, and from the lack of homoserine in the amino-acid composition of peptide CNBr-2, it is probable that peptide CNBr-2 is derived from the *C*-terminal portion of factor D̄ (Fig. 4). Peptide CNBr-2 was estimated to be composed of approx. 60 amino acid residues, and automated *N*-terminal sequence analysis gave sequence of the first 26 amino acid residues (with the exception of positions 21 and 22). When peptide CNBr-2 was isolated from a CNBr digest of factor D̄ that had been labelled with [1,3-^{14}C]Pr_2P-F, it was found to be radiolabelled. By scintillation counting of one-third of the butyl chloride extracts from the sequencer run of peptide CNBr-2 it was shown that the radioactivity in the peptide was located in the serine residue found at position 14 (Table 3).

Discussion

The purification procedure described in this paper gave a final yield of about 0.2mg of factor D̄/1000ml of serum, and a recovery of about 20% of the initial factor D̄ haemolytic activity (Table 1). The overall purification from serum was approx. 60000-fold. Addition of protease inhibitors in the early stages of the purification did not appear to increase the recovery of factor D̄. The yield of factor D̄ is consistent with a serum concentration of approx. 1.0–1.5μg/ml, which is similar to the estimates of Lesavre *et al.* (1979) but slightly lower than suggested by an earlier report (Lesavre & Müller-Eberhard, 1978).

The estimate of an apparent mol.wt. of 24000 is similar to values obtained by others (Volanakis *et al.*, 1977). Factor D̄ that had been reduced and alkylated showed a slightly slower mobility on SDS/polyacrylamide gels compared with non-reduced factor D̄ (Fig. 2). This has also been observed by Volanakis *et al.* (1977). Lesavre *et al.* (1979) and Davis *et al.* (1979*b*) found that factor D̄ stained with Schiff stain after periodate oxidation, which indicates that it is a glycoprotein. In the

Table 3. *Automated sequence analysis of factor D̄ and peptides CNBr-1, CNBr-2 and CNBr-3*

Details of the automated Edman degradation are given in the text. Amino-acid phenylthiohydantoin derivatives were identified by high pressure liquid chromatography. The presence of the *S*-carboxymethylcysteine phenylthiohydantoin derivatives were also confirmed by detection of radioactive label. —, not identified.

	Amino acid identified in			
Residue number	Factor D̄	Peptide CNBr-1	Peptide CNBr-2	Peptide CNBr-3
1	Ile	Ala	Cys	Ile
2	Leu	Ser	Ala	Leu
3	Gly	Val	Glu	Gly
4	Gly	Gln	Ser	Gly
5	Arg	Leu	Asn	Arg
6	Glu	Asn	Arg	Glu
7	Ala	—	Arg	Ala
8	Glu	Ala	Asp	Glu
9	Ala	—	Ser	Ala
10	His	Leu	Cys	—
11	Ala	Cys	Lys	Ala
12	Arg	Gly	Gly	[illegible]
13	Pro	Gly	Asp	
14	Tyr	Val	Ser*	
15	Met	Leu	Gly	
16	Ala	Val	Gly	
17	Ser	Ala	Pro	
18	Val	Glu	Leu	
19	Gln	—	Val	
20	Leu	—	Cys	
21	Asn	Val	—	
22	—	[illegible]	—	
23	Ala		Val	
24	Glu		Leu	
25	Leu		Glu	
26	—		(Gly)	
27	Gly			
28	Gly			
29	Val			
30	Leu			
31	Val			
32	Ala			
33	Glu			
34	Gln			
35	—			
36	Val			
Amount used in sequencer (nmol)	60	20	30	12
Recovery at first step (nmol)	6.5	3.1	10.5	3.1
Stepwise yield (%)	93	90	94	75
Residues used to calculate yield	Leu-2,20,25,30	Leu-5,10,15	Cys-1,10,20	Ala-7,9,11

* Radioactivity was released at this position when peptide CNBr-2 was isolated from [1,3-^{14}C]iPr_2P-F-labelled factor D̄.

purification procedure used in this study, however, it was observed that factor D̄ was not retained by a concanavalin A–Sepharose column.

The amino-acid composition obtained for factor D̄ shows no unusual features and is in good agreement with that reported by Davis *et al.* (1979*b*) (Table 2). However, the one other published amino-acid composition given for human factor D̄ (Volanakis *et al.*, 1977) differs in some aspects from the values given in this study (Table 2), e.g. isoleucine was reported to be present as 6.0 residues/100 residues compared with 2.6 residues/100 residues found in this study; alanine 5.8 compared with 8.8; cystine 2.9 compared with 4.0; leucine 6.9 compared with 10.3; lysine 7.1 compared with 3.9. The methionine value of 1.2 residues/100 residues

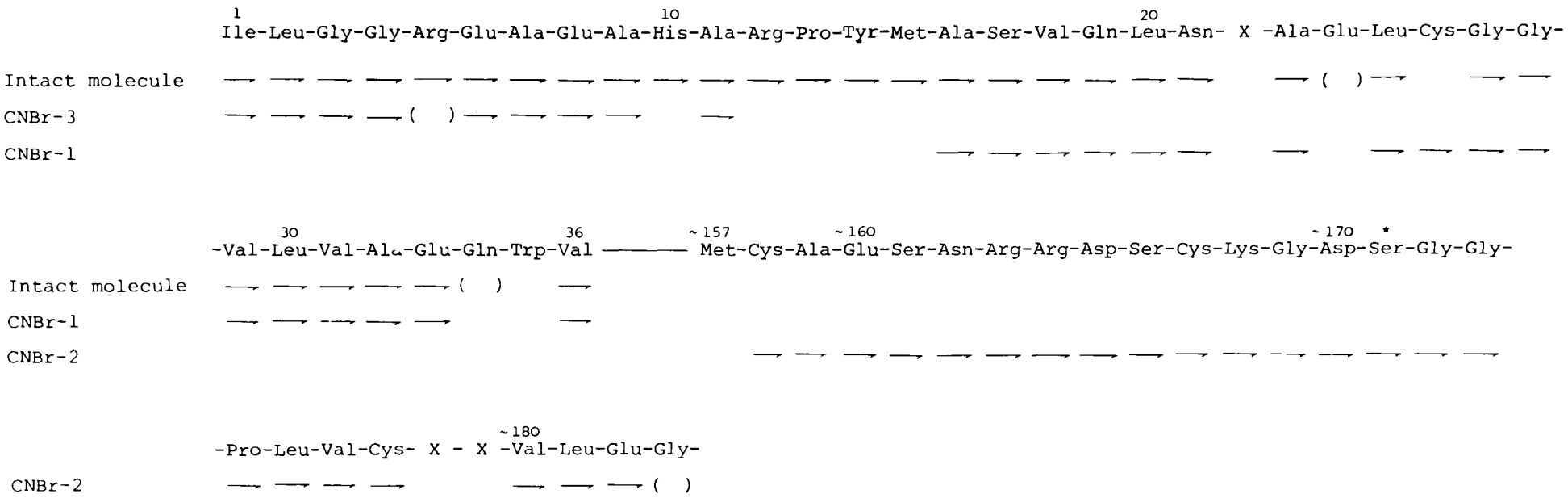


Fig. 4. *N-Terminal amino-acid sequence of factor D̄ and alignment of the major CNBr fragments*

The intact molecule, peptide CNBr-1, peptide CNBr-2 and peptide CNBr-3 have apparent mol.wts. of 24000, 15800, 6600 and 1700 respectively and therefore are composed of approx. 215, 141, 60 and 15 amino acid residues respectively. —, ( ) or a blank space denote that positive identification, tentative identification or no identification respectively of a particular residue was made. * Denotes the position of the [1,3-^{14}C]iPr_2P-F-labelled serine residue. The residue at position 35 was identified as tryptophan by examination of the peptides produced by the selective cleavage of the tryptophanyl bonds present in factor D̄ (D. M. A. Johnson, J. Gagnon and K. B. M. Reid, unpublished work).

obtained in this study (Table 2) indicated that there were 2–3 methionine residues/molecule of factor $\bar{D}$, which should therefore yield three or four peptides on treatment with CNBr.

Digestion of factor $\bar{D}$ with CNBr gave three major peptides, CNBr-1, CNBr-2 and CNBr-3 (Table 2 and Fig. 4) in yields of 26%, 55% and 39% respectively. The yields of peptides CNBr-1 and CNBr-3 were lower than that of peptide CNBr-2 since they each underwent an extra purification step compared with peptide CNBr-2. The amino-acid-sequence data given allows the alignment of these CNBr peptides as shown in Fig. 4. The possibility that there may be a fourth, very small, peptide has not been eliminated and must be considered, especially as it can be estimated that intact factor $\bar{D}$ contains approx. 2.5 residues of methionine/molecule (Table 2). That any unidentified peptide must be small is suggested from the finding that the apparent molecular weight of factor $\bar{D}$ is virtually accounted for by the sum of the molecular weights of peptides CNBr-1, CNBr-2 and CNBr-3 (15800, 6600 and 1700 respectively).

The yield of *N*-terminal phenylthiohydantoin amino acid at the first cycle in each of the sequencer runs performed on factor $\bar{D}$ and the CNBr fragments was usually low (Table 3). The reasons for this low yield in most of the sequencer runs are not known, but could involve losses of the polypeptide from the cup during the washing steps in the first cycle of the sequencer programme. The possibility that general technical difficulties caused the low yields can be excluded since myoglobin standards, run on the sequencer using the 0.3 M-Quadrol programme, consistently gave a yield of at least 70% of that expected for the recovery of the *N*-terminal phenylthiohydantoin amino acid. Since the yield on the first cycle on each of the sequencer runs was low this made the interpretation of the sequence data, beyond cycles 36 and 26, difficult for the sequencer runs on factor $\bar{D}$ and peptide CNBr-2 and consequently no data beyond cycles 36 and 26, in these two runs, has been reported (Table 3). If unequivocal identification of a residue was not made this has been stated in the results. The phenylthiohydantoin amino acids released were readily detected in amounts of 0.5 nmol in the sequencer runs and no problems were encountered with spurious background peaks or 'carry-over' of phenylthiohydantoin amino acids. The phenylthiohydantoin derivatives of tryptophan residues are normally readily identified by high pressure liquid chromatography and thus it is unlikely that the unidentified residue at position 22 (Figs. 4 and 5*a*) is tryptophan. The tryptophan at position 35 was identified by examination of the peptides produced by the selective cleavage of the tryptophanyl bonds present in factor $\bar{D}$ (D. M. A. Johnson, J. Gagnon and K. B. M. Reid, unpublished work). The reason why positive identification of this residue was not made from the sequencer runs reported in this paper is that position 35 is very close to the point in the runs where it was considered that unequivocal identification of the phenylthiohydantoin derivatives could not be made.

Davis *et al.* (1979*a*,*b*) have reported that the *N*-terminal amino-acid sequence of factor $\bar{D}$ is:

Ile-Leu-Gly-Gly-Arg-Glx-
(Ala or Ser)-Glx-(Ala or Ser)

The *N*-terminal amino-acid sequence reported in this study agrees exactly with the identification of positions 1–5 reported by Davis *et al.* (1979*a*,*b*) and also is in agreement with the identification of positions 6–9 reported by Davis *et al.* (1979*a*,*b*) who were unable to distinguish between glutamic acid and glutamine and between alanine and serine in the procedure they used for identification of the amino-acid phenylthiohydantoin derivatives.

The *N*-terminal amino-acid sequence of factor D is shown aligned with the *N*-terminal amino acid sequences of the catalytic chains of other serine esterases in Fig. 5(*a*). Using the method of Moore and Goodman (1977) a preliminary statistical evaluation of the similarities between factor $\bar{D}$, subcomponents $C\bar{1}r$, $C\bar{1}s$, 'group-specific protease', elastase, chymotrypsin, trypsin, thrombin and factor X can be made. The method is based on the number of residues aligned and the minimum number of mutations required to make the sequences identical. Over the first 20 residues (which is all that has been published of the *N*-terminal amino acid sequences of the catalytic chains of subcomponents $C\bar{1}$ and $C\bar{1}s$) the probability of a random relationship for factor $\bar{D}$ versus rat 'group-specific protease', elastase and subcomponent $C\bar{1}r$ or $C\bar{1}s$ was 0.01, 0.01 and 1% respectively. Comparison of subcomponent $C\bar{1}r$ with subcomponent $C\bar{1}s$ gave a value of 0.01%. The significance levels obtained on comparison (Fig. 5*a*) of factor $\bar{D}$ with trypsin, chymotrypsin, plasmin, thrombin and factor X, over the first 20 residues, were all in the range $>0.1\% \leqslant 1\%$. Thus the *N*-terminal amino acid sequence of factor $\bar{D}$ appears more similar to that of rat 'group-specific protease', or elastase, than to the *N*-terminal sequence of the complement subcomponents $C\bar{1}r$ and $C\bar{1}s$. This apparent strong homology of factor D with the 'group-specific protease' from rat small intestine is of interest since it is considered that the 'group-specific protease' may have no zymogen form (Woodbury *et al.*, 1978) while it is unknown whether factor $\bar{D}$ has a zymogen form or not. It can also be seen from Fig. 5(*a*) that the strictly conserved residues in the 1–36 region of the *N*-terminal sequences of the catalytic chains of other serine esterases are also conserved in factor $\bar{D}$, i.e. glycine at position 4, proline at 13, cystine at 26, glycine at 27 and valine at 36. The

(*a*)

	1									10										20										30						36
Factor D̄	I	L	G	G	R	E	A	E	A	H	A	R	P	Y	M	A	S	V	Q	L	-	-	-	-	N	X	A	E	L	C	G	G	V	L	V	A	E	Q	W	V
'Group-specific protease' (rat)	I	I	G	G	V	E	S	I	P	H	S	R	P	Y	M	A	H	L	D	I	V	T	E	K	G	L	R	V	I	C	G	G	F	L	I	S	R	Q	F	V
Elastase (pig)	V	V	G	G	T	E	A	Q	R	N	S	W	P	S	Q	I	S	L	Q	Y	R	S	G	S	S	W	A	H	T	C	G	G	T	L	I	R	Q	N	W	V
C1̄r (human) 'b' chain	I	I	G	G	Q	K	A	K	M	G	N	F	P	W	Q	V	F	T	N	Z																				
C1̄s (human) 'b' chain	I	I	G	G	S	D	A	D	I	K	N	F	P	W	Q	V	F	F	D	N																				
Chymotrypsin A (cow)	I	V	N	G	E	E	A	V	P	G	S	W	P	W	Q	V	S	L	Q	D	K	T	G	F	-	-	-	H	F	C	G	G	S	L	I	N	E	N	W	V
Plasmin (human)	V	V	G	G	C	V	A	H	P	H	S	W	P	W	Q	V	S	L	R	T	R	F	G	M	-	-	-	H	F	C	G	G	T	L	I	S	P	E	W	V
Trypsin (cow)	I	V	G	G	Y	T	C	G	A	N	T	V	P	Y	Q	V	S	L	-	-	N	S	G	Y	-	-	-	H	F	C	G	G	S	L	I	N	S	Q	W	V
Thrombin (cow)	I	V	E	G	Q	D	A	E	V	G	L	S	P	W	Q	V	M	L	F	R	K	S	P	Q	E	-	-	L	L	C	G	A	S	L	I	S	D	R	W	V
Factor X (cow)	I	V	G	G	R	D	C	A	E	G	E	C	P	W	Q	A	L	L	V	-	N	E	E	N	E	-	-	G	F	C	G	G	T	I	L	N	E	F	Y	V

(*b*)

	181																190																200		
Factor D̄	M	C	A	E	S	-	-	-	-	N	R	R	-	D	S	C	K	G	D	S	G	G	P	L	V	C									
'Group-specific protease' (rat)	V	C	V	G	S	-	-	P	T	T	L	R	-	A	A	F	M	G	D	S	G	G	P	L	L	C									
Elastase (pig)	V	C	A	G	-	-	-	G	N	G	V	R	-	S	G	C	Q	G	D	S	G	G	P	L	H	C									
Chymotrypsin A (cow)	I	C	A	G	-	-	-	A	S	G	V	-	-	S	S	C	M	G	D	S	G	G	P	L	V	C									
Plasmin (human)	L	C	A	G	H	-	-	L	A	G	G	T	-	D	S	C	Q	G	D	S	G	G	P	L	V	C									
Trypsin (cow)	F	C	A	G	Y	-	-	L	E	G	G	K	-	D	S	C	Q	G	D	S	G	G	P	V	V	C									
Thrombin (cow)	F	C	A	G	Y	K	P	G	E	G	K	R	G	D	A	C	E	G	D	S	G	G	P	F	V	M									
Factor X (cow)	F	C	A	G	Y	-	P	D	T	Q	P	E	-	D	A	C	Q	G	D	S	G	G	P	H	V	T									

Fig. 5. *Alignment of* (*a*) *the N-terminal region and* (*b*) *the active site of factor* D̄ *with N-terminal regions and active sites of other serine esterases*
Sequence data for subcomponents C1̄r and C1̄s are from Sim *et al.* (1977). The sequences given for 'group-specific protease' are from Woodburg *et al.* (1978). Sequence data for chymotrypsin, trypsin, elastase, thrombin and plasmin are from de Haën *et al.* (1975) and Dayhoff (1978). *a*, *N*-Terminal sequence; the residue numbering shown is that of the factor D̄ sequence. *b*, Sequence around the active-site serine residue; the residue numbering shown is that of the cow chymotrypsin A sequence. ——, Denotes that a gap was left to give maximum homology on alignment of the amino-acid residues is used: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, unknown; Y, Tyr.

presence of two hydrophobic residues at the *N*-terminus of the molecule has also been retained, which is consistent with the probable importance of these residues in the formation of the substrate binding site (Kraut, 1971).

In most serine proteases the active site serine is located approx. 46 amino-acid residues from the *C*-terminal end of the molecule (de Haën *et al.*, 1975). The peptide CNBr-2 is approx. 60 amino acid residues long and is located in the *C*-terminal region of the molecule (Fig. 3). This peptide was found to be labelled with radioactivity after its isolation from a CNBr digest of factor $\bar{D}$ which had been previously labelled with [1,3-^{14}C]iPr_2P-F. The radioactive label was found to be located on the serine at position 14 of peptide CNBr-2 (Table 3) and the amino-acid sequence around this position is very typical of the highly conserved amino-acid sequence found around the active site serines of other serine esterases (Fig. 5*b*). It is probable that the relevant Met–Cys bond that was split to yield peptide CNBr-2 corresponds to the Ile_{181}–Cys_{182} bond of bovine chymotrypsin A (Fig. 5*b*). In dogfish trypsin the equivalent portion of amino acid sequence is -Met-Met-Cys- (de Haën *et al.*, 1975).

Factor $\bar{D}$ splits factor B at a single Arg–Lys bond, thus indicating a trypsin-like specificity (Lesavre *et al.*, 1979). The presence of aspartic acid at position 8 in peptide CNBr-2 (Fig. 3) is consistent with factor $\bar{D}$ belonging to a family of trypsin-like enzymes all of which have Asp_{189} (Fig. 5*b*) (chymotrypsin numbering) present in the substrate-binding pocket.

We thank Mr. T. Gascoyne for performing the amino-acid analyses and Mr. A. C. Willis and Mrs J. Parsons for excellent technical assistance. D. M. A. J. held a Commonwealth Scholarship from the Association of Commonwealth Universities.

References

Austen, K. F. (1978) *J. Immunol.* **121**, 793–805
Bridgen, P. J., Cross, G. A. M. & Bridgen, J. (1976) *Nature (London)* **263**, 613–614
Davis, A. E., Zalut, C., Rosen, F. S. & Alper, C. A. (1979*a*) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* (Abstract) **38**, 1216
Davis, A. E., Zalut, C., Rosen, F. S. & Alper, C. A. (1979*b*) *Biochemistry* **18**, 5082–5087
Dayhoff, M. O. (ed.) (1978) *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, National Biomedical Research Foundations, Silver Springs
de Haën, C., Neurath, H. & Teller, D. C. (1975) *J. Mol. Biol.* **92**, 225–259
Fearon, D. T., Austen, K. F. & Ruddy, S. (1974) *J. Exp. Med.* **139**, 355–366
Götze, O. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* (Abstract) **35**, 269
Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133
Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131
Kraut, J. (1971) *Enzymes 3rd Ed.* **3**, 165–185
Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
Lesavre, P. H. & Müller-Eberhard, H. J. (1978) *J. Exp. Med.* **148**, 1498–1509
Lesavre, P. H., Hugli, T. E., Esser, A. F. & Müller-Eberhard, H. J. (1979) *J. Immunol.* **123**, 529–534
Martin, A., Lachmann, P. J., Halbwachs, L. & Hobart, M. J. (1976) *Immunochemistry* **13**, 317–324
Moore, G. W. & Goodman, M. (1977) *J. Mol. Evol.* **9**, 121–130
Schreiber, R. D., Pangburn, M. K., Lesavre, P. H. & Müller-Eberhard, H. J. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3948–3952
Sim, R. B., Porter, R. R., Reid, K. B. M. & Gigli, I. (1977) *Biochem. J.* **163**, 219–227
Volanakis, J. E., Schrohenloher, R. E. & Stroud, R. M. (1977) *J. Immunol.* **119**, 337–342
Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K. & Neurath, H. (1978) *Biochemistry* **17**, 811–819