STUDIES

ON

IMMUNOGLOBULIN VARIABILITY

by

David Stanley Secher

A disstertation submitted to the University of Cambridge in candidature for the degree of Doctor of Philosophy

August 1974

Gonville and Caius College
Cambridge

PREFACE

The work described in this dissertation was carried out in the Laboratory of Molecular Biology, Cambridge, between 1971 and 1974 during the tenure of a Medical Research Council Scholarship for Training in Research Methods and a Fellowship from the Salters' Company, London. All the work described is my own, except where otherwise stated in the text.

I should like to thank Dr F. Sanger, head of the Division of Protein and Nucleic Acid Chemistry, and other colleagues, in particular Drs N.J. Cowan, R.G.H. Cotton and J. Bridgen, for their helpful advice and discussion. I am especially grateful to my supervisor, Dr C. Milstein, for his constant and enthusiastic encouragement.

Fred Northrop and John Jarvis provided useful practical help in the laboratory, and Shirley Howe helped with the care of cell cultures. Dave Hart and the late Len Hayward constructed isoelectric focusing apparatus. Lynne Toward typed the dissertation with great patience and accuracy and Annette Lenton, Ken Harvey and Judith Pryor provided expert assistance with the figures and photographs.

My thanks to all these and also to Peggy Dowding.

This dissertation is not substantially the same as any submitted for any degree or other qualification at any other university and no part of it has been, or is being concurrently, submitted for any such degree, diploma, or other qualification.

David Secher

CONTENTS

CHAPTER 1

Introduction

CHAPTER 1

Introduction

The ability of each individual to recognise with exquisite specificity a seemingly infinite range of foreign substances is one of the most challenging observations of molecular immunology. The mechanism by which the immune system is capable of evolving or generating this enormous diversity remains unknown. This dissertation describes two independant experimental approaches to understanding the variety of antibodies. The first extends the amino acid sequence data to immunoglobulins of a species not previously investigated; the second involves a study of spontaneous mutation in a model tissue culture system.

Accordingly this chapter consists of a section on some aspects of the structure of antibodies relating to sequence data, followed by a section comparing the hypotheses proposed to explain the observed variation.

1.1 Structure of Antibodies

This topic has been extensively reviewed recently (Porter, 1973; Gally, 1973; Smith, 1973, Milstein & Pink, 1970; Smith et al, 1971) and it is therefore only necessary to summarise the principal conclusions here.

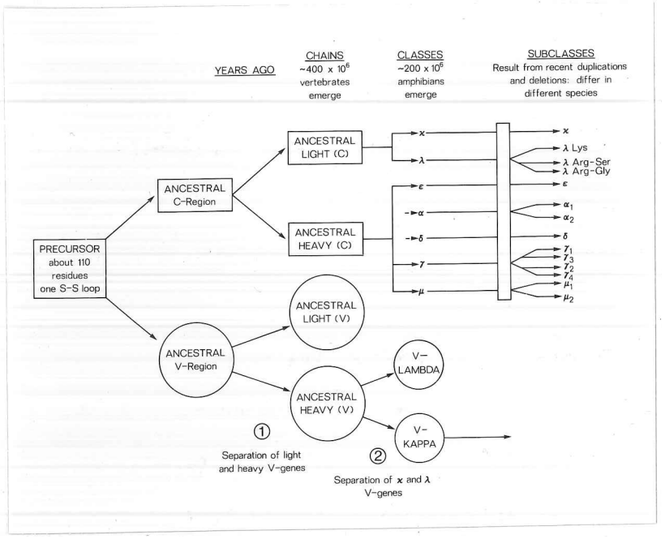
Antibodies are immunoglobulins (Ig's), the basic unit of which consists of two identical "Heavy" (H-) and two identical "Light" (L-) chains. Some antibodies consist of a largely homogeneous population of molecules (e.g. some rabbit anti-polysaccharide antibodies, Krause, 1970); more usually they are a heterogeneous population and complete amino acid sequence studies are impossible. Most of the data on Ig sequences has therefore been derived from myeloma proteins, found in the serum of man

and several other mammals as a result of a variety of lymphoproliferative disorders. In general these proteins are homogeneous Ig's and it is generally accepted that they arise from the random proliferation of a particular Ig-producing cell. Thus the conclusions of studies about structure and variety of myeloma proteins are assumed to be applicable to antibodies in normal individuals.

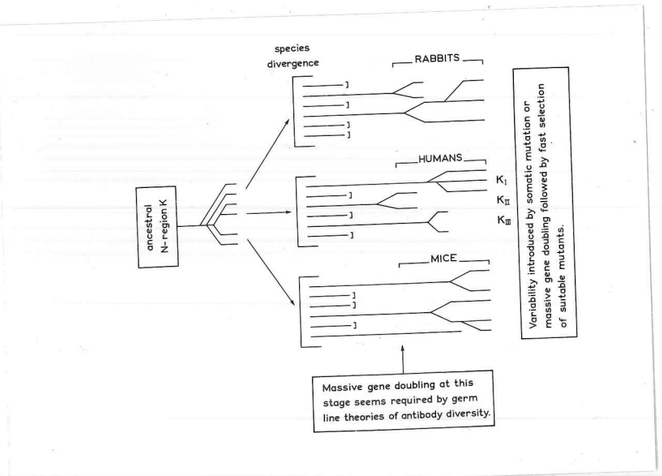
Schwarz and Edelman (1963) showed that the Bence-Jones proteins excreted in the urine of some human myeloma patients are identical to the L-chains of their myeloma proteins. These L-chains were the first Ig chains whose sequences were determined (Hilschmann & Craig, 1965; Titani & Putman, 1965). Subsequently the sequences of many human L-chains as well as a human H-chain (Edelman et al, 1969) have been determined. Tumours producing myeloma proteins may be induced in Balb/c mice and have also been studied in detail (reviewed in Potter, 1972). Complete or near complete sequences have been reported for L-chains (Svasti & Milstein, 1972b; McKean et al, 1973; Gray et al, 1967) and for H-chains (Milstein et al, 1974b). In other species partial sequences of H- and L-chains have also been obtained either from myeloma proteins or from normal antibodies of naturally restricted heterogeneity. The available sequences have recently been collated by Smith (1973).

1.1.1 Ig-chains

L-chains contain about 214 residues comprising two homologous "pseudo-subunits" each of about 107 residues, containing a single disulphide-bridged loop (Milstein, 1964; Hilschmann & Craig, 1965). The amino-terminal "V-region" differs between different L-chains, whereas each carboxy-terminal "C-region" is common to many L-chains and defines the type (κ or λ) and sub-type (see below). Similarly H-chains also contain V- and C-regions: in this case the V-region consists of one pseudo-subunit, the C-region of three or four (C_{H1}, C_{H2}, etc.).



a



b

Fig. 1.1 Possible scheme for the evolution of immunoglobulins. (1) and (2) indicate stages of evolution at which duplication of whole chromosomes has occurred.

a) shows scheme for evolution of classes, sub-classes and types.

b) is a continuation of part of (a) and illustrates a pathway of evolution for V_K genes which could explain the general similarities and specific differences between sub-groups in different species (Milstein and Pink, 1970).

The observation of internal homology in L-chains and in H-chains led to the suggestion that they evolved by gene duplication of an ancestral gene which had a single disulphide-bridged loop (Hill et al, 1966). A modern scheme to explain the observed homologies in human Ig's is shown in Fig. 1.1.

The sequence homology of the ψ-subunits, together with data from electron microscopy (Valentine & Green, 1967; Feinstein & Munn, 1969) and proteolytic digestion (Porter, 1959; Nisonoff et al, 1960; Turner & Bennich, 1968; Kehoe & Fougereau, 1969) suggested that there might also be three-dimensional structural homology. This has recently been confirmed crystallographically for the antigen-binding fragment of a human myeloma protein (Poljak et al, 1973) and for a L-chain dimer (Schiffer et al, 1973).

1.1.2 C-region Variation

Within a species the greatest C-region variation is between the different classes of Ig H-chain. These classes, which are also distinguishable on the basis of physico-chemical and antigenic properties are present in all individuals of a species and therefore the products of separate gene loci. Humans, for example, produce 5 classes of H-chain: γ, μ, α, δ, ε. When combined with a κ or λ L-chain these H-chains give rise to IgG, IgM, IgA, IgD and IgE respectively. At least some of the classes of H-chains may be further divided, on the basis of their biological properties, antigenic structure, or amino acid sequence, into subclasses which are also present in all individuals.

A similar division exists in human λ chains where a Lys/Arg interchange at position 191 and a Ser/Gly interchange at 153 define three subtypes that presumably arose following recent gene duplications (Milstein, 1967; Appella & Ein, 1967; Hess et al, 1971).

Allotypic variants (i.e. polymorphic forms of a single locus) have also been found in C-region genes (reviewed in Mage et al, 1973; Milstein & Munro, 1973; 1970). These genetic markers, which have been correlated in some instances with changes in the amino acid sequence (Milstein, 1966a; Baglioni et al, 1966), support the conclusion that C-genes exist (or at least behave) as single copies. Linkage studies using allotypic markers show that the C-genes for (rabbit) κ,λ and H-chains are unlinked to each other (Mage, 1971) but that the (mouse) H-chain genes are closely linked (Potter and Lieberman, 1967).

1.1.3 V-region Variation

The variation observed in V-region sequences is much more extensive. Identical V-region sequences on different Ig's are seldom detected (Quattrochi et al, 1969; Capra & Kunkel, 1970). The variation is not however distributed over the whole V-region equally. Certain residues are invariant (for example the half-cystines) or almost so. It seems likely that this invariance is dictated by structural requirements of the molecule for folding, chain recognition, secretion, etc. Other residues vary only between different "families" (κ,λ,H) or species of V-regions. Within such groups, "sub-groups" may in some cases be defined from the sequence data and sub-group specific residues identified.

1.1.4 Hypervariable Residues

Three sections of L-chains and of H-chains are characterised by an extremely high rate of amino acid substitutions and insertions between different V-regions (Wu & Kabat, 1970; Kabat & Wu, 1972). These "hypervariable" sections occur around positions 28-34, 50-54, 80-96 in L-chains and at 30-35, 50-52 and 95-100 in H-chains. Kehoe and Capra

(1971) have suggested a further hypervariable region in position 86-91 but, in the rabbit, Mole et al (1971) found this region to be correlated with allotype.

The high degree of variability suggests that these residues confer the antigen-binding specificity on the Ig. This conclusion was supported by affinity labelling studies (Singer & Doolittle, 1966; Franék, 1971; Fleet et al, 1972) and recently it has been shown crystallographically that these sections are in contact with the antigen in the case of the human myeloma protein NEW (Poljak et al, 1973).

1.1.5 Association of chains

Variation is also possible, at least in theory, at the level of H- and L-chain combination. It has been reported that in a competition experiment L-chains preferentially reassociate with the H-chains from which they were separated (Stevenson & Mole, 1974). However the principle of random association of H- and L-chains ("pxq hypothesis") is often assumed in calculations of antibody diversity (Edelman & Gally, 1968; Cunningham et al, 1971).

1.1.6 "Abnormal" Variants

The above variation observed in myeloma proteins is believed to reflect the natural variation of Ig's. In addition to these are certain rare variants which are presumed to arise from "abnormal" genetic events. Such mutants may however be very useful in the understanding of the normal situation.

An example of this type of variant is the "Lepore"-type hybrid $\gamma 3$-$\gamma 1$ chain (Kunkel et al, 1969) which was considered to result from a non-homologous crossing-over between the $\gamma 1$ and $\gamma 3$ cistrons. Since $\gamma 2$ and $\gamma 4$ are still expressed $\gamma 3$ and $\gamma 1$ are presumably adjacent in the C_H-

gene cluster.

The Heavy Chain Disease proteins are another type of naturally occurring mutant Ig. These proteins, found in the sera of certain myeloma patients, appear to be derived from H-chains by the deletion of a number of consecutive residues (up to about 200). In those cases where primary structure data are available, residue 216 (Eu numbering) is often involved at one end of the deletion (see Frangione & Franklin, 1973 for review). The significance of the repeated involvement of this area is not clear. It has been suggested that it represents the initiation of a second C-gene (Franklin & Frangione, 1971) but this is not supported by any genetic evidence. An alternative explanation has been proposed by Milstein et al (1974a) who suggest, on the basis of the Heavy Chain Disease proteins and also IF-2 (Chapter 6) that the deletions represent incorrect integration of V- and C-regions and that residue 216 is close to, or part of, a region homologous to the V-C integration signal and derived from it by tandem duplication of an ancestral gene.

A variant mouse IgA was discovered by Potter et al (1964) and shown by Mushinski (1971) to be a shortened H-chain. This could also be a similar class of proteins to the Heavy Chain Disease proteins but confirmation of this will require more detailed chemical analysis now in progress (Robinson, E.A., Smith, D.F. & Appella, E., personal communication).

1.2 Genetic Basis of Variation

The existence of a single C-region common to many Ig's, each defined by a different V-region (Milstein, 1966b; Hilschmann & Craig, 1965) implies either that the V- and C-regions are defined by separate V-genes and C-genes (Dreyer & Bennett, 1965) or that there are multiple copies of each C-gene (Brown, 1972). Mechanisms by which such multiple

Man.

		V-genes	C-genes
Light Chains	κ	Ia Ib II III	—
	λ	I II III IV	Arg Lys Gly
Heavy Chains		I II III	γ4 γ2 γ3 γ1 α1 α2 μ2 μ1 δ ε

Mouse.

		V-genes	C-genes
Light Chains	κ	I II III IV V VI VII VIII etc.	—
	λ	I [II]	I II

Fig. 1.2 Possible arrangement of the minimum number of genes for human and mouse light chains. The genes on each horizontal line are thought to lie on the same chromosome. In mice the minimum number of V_κ-genes is probably well above 8 and the existence of a second V_λ-gene (Schulenberg et al, 1971) is uncertain (Milstein & Munro, 1973).

copies might be maintained constant and free of mutation have been proposed (Callan, 1967; Whitehouse, 1973) but most immunologists now agree that there is a single gene for each C-region as shown in Fig. 1.2. The number of V-genes is a topic of considerable dispute (see below) but is at least equal to the number of sub-groups where these may be clearly defined. This assumes that the sub-groups are not allelic forms as has been demonstrated for κ_{Ib} and κ_{II} (Milstein et al, 1969) and for H-chains (Wang et al, 1971) of humans. The V_κ sequences are always associated with C_κ and likewise V_λ with C_λ. There appears to be no preferential association of any of the V_λ sub-groups with one of the sub-types (Ein & Hood, 1968; Hess et al, 1971). These observations are consistent with the existance of two families of closely linked genes, probably on separate chromosomes. The H-chain genes form a third family unlinked to the other two, with the V_H genes shared amongst the classes and subclasses of C_H genes. The sharing of V_H genes was first suggested following the observation that rabbit V-region allotypes are found in γ, μ and α chains. (Todd, 1963; Feinstein et al, 1963; Todd & Inman, 1967). A similar result has since been shown for human H-chains (Pink et al, 1972).

1.2.1 One cell - one Ig

The theory that each Ig-producing cell synthesises a single species of Ig - a fundamental tenet of the now generally accepted clonal selection theory (Burnet, 1959; Jerne, 1955) - has received further support recently from both in vivo (Askonas et al, 1970) and in vitro (Awdeh et al, 1970) studies. There are however exceptions to the principle which suggest that the differentiated plasma cell is capable of undergoing further mutational events, especially retranslocation of the V-gene. Firstly the observation of single plasma cells secreting IgG and IgM molecules of the same specificity (Nossal et al, 1964) and the similar finding of the same idotypes on IgM early in an immune response and on

IgG later (Oudin & Michel, 1969). That the same phenomenon occurs in man is suggested from the study of a rare myeloma case in which two clones of cells proliferate, one producing IgG, the other IgM (Levin et al, 1971). The L-chains and V_H regions appear to be identical in the two proteins.

These results suggest that translocation of a V-gene to a new C-gene is possible and thus a likely mechanism for the joining of V- and C-region information. The Heavy Chain Disease proteins and IF-2 (see chapter 6) may represent abnormal examples of this translocation (or "retranslocation") and thus provide a clue to the mechanism of these apparent changes in the genome. (For the evidence that the joining is unlikely to be at the level of RNA or protein see Gally, 1973; Milstein et al, 1974a; Milstein & Munro, 1970.)

A second type of cell producing two Ig's was recently reported by Cotton & Milstein (1973). Fusion with Sendai virus of two Ig-secreting myeloma cells (one rat line and one mouse line) resulted in the successful isolation for the first time of a hybrid cell which continued to secrete both parental Ig's. The observation that mixed molecules e.g. mouse H-chains attached to rat L-chains but not hybrid chains i.e. mouse V-regions attached to rat C-regions or vice versa is further evidence that the integration of information for V- and C-regions takes place in the genome and does not involve diffusible gene products.

1.2.2 How many germ-line V-genes?

Proponents of germ-line theories (Hood & Talmage, 1970; Hilschmann et al, 1969; Dreyer et al, 1967) maintain that the genes for all the immunoglobulins that an individual can produce are carried in the germ-line and that the mechanisms for expressing a single V-C

ANTIBODY SPECIES RAISED AGAINST HAPTEN A

No.	1	2	3	4	5 ..... h
Specificities	A	A	A	A	A A
	A'	A'	A'		A' A'
	A"		A"	A"	A"
	B	C	F	J	(D) T
	E	(D)	G	W	H L
	Q	M	R	X	P Y
	V	S	N	K	Z U
	⋮	⋮	⋮	⋮	⋮ ⋮

Fig. 1.3 Specificity profiles of individual antibody species in a hypothetical antiserum illustrating the concepts of multispecificity. The letters stand for determinant-sized haptens which, in each vertical column, represent the potential reactivities or specificities of an individual antibody molecular species. A' and A" are closely related in structure to the immunogen hapten A (A' more so than A"). The other haptens are, for the most part, structurally unrelated to A or to one another. The disparate reactivities are largely unshared between different antibody species. Species numbers 2 and 5 share D-reactivity and may provide for a weakly positive serological reaction to hapten D in a suitable test system (Inman, 1974).

combination per cell is analagous to and an extreme case of the expression of characteristic products of differentiated cells.

Somatic theories conversely postulate the existence of a small number of V-genes (e.g. not less than 4 in the case of human V_κ) upon which somatic processes operate to create diversity. Somatic mutation (Cohn, 1968; Jerne, 1971), hypermutation (Brenner & Milstein, 1966) and various forms of recombination (Smithies, 1967; 1970; Gally and Edelman, 1970) have been proposed as possible mechanisms.

Recent additions to the large volume of sequence data have not clarified the data in favour of one type of theory. The observation that rabbit V_H (a-locus) allotypes segregate as Mendelien alleles argues in favour of a small number of genes and hence somatic diversification (Kelus & Gell, 1967). Conversely the sequences of mouse κ-chain N-termini (Hood et al, 1973) do not fit into a small number of well-defined sub-groups and thus even a somatic theory for mouse κ-chain diversity requires, say, 30-100 V-genes.

1.2.3 How much diversity is required?

It is often assumed that the repertoire of immunological responses reflects, on a one-to-one basis, the number of different antibodies produced i.e. that each antibody has a unique antigen. The initial observations of Landsteiner (1945), demonstrating a high degree of specificity of hapten recognition, referred to heterogeneous antibody populations. It has been noted, but often disregarded, that such specificity might alternatively arise from the cumulative effect of a large number of multiple specificity antibodies (Fig. 1.3) (Talmage, 1959; Williamson, 1972; Eisen, 1971; Milstein & Munro, 1973; Inman, 1974). The recent observation that myeloma proteins (Varga and Richards, personal communication quoted in Inman, 1974) and antibodies (Varga et al, 1973) exhibit cross-reactivity with apparently unrelated antigens has given more weight to these arguments, which could reduce

considerably the number of genes required by a germ-line theory.

1.2.4 Predictions

To date the amino acid sequence data have failed to distinguish even the two general types of theory and although it is believed by some immunologists that the resolution of the problem will come only from a knowledge of more sequences, others believe that a closer examination of different predictions of the hypotheses is required. For example, "Probably most scientists who have an interest in this topic are weary of the theoretical arguments, which are beginning to become a bit stale by now, and look forward to fresh experimental data to resolve the question", (Gally, 1973).

1.2.5 Nucleic acid hybridisation

Germ-line theories require a large number of V-genes (several per sub-group) whereas somatic theories predict that a single Ig-producing cell (or a germ cell) will only contain one gene per sub-group. Now that mRNA coding for Ig L-chains is available in a pure or nearly pure state (Brownlee et al, 1973; Swan et al, 1972; Stevens & Williamson, 1972; Mach et al, 1973; Tonegawa et al, 1974) it is theoretically possible to measure the reiteration frequency of those regions of the DNA which hybridise to the mRNA (or a complementary DNA) probe. There are still problems in the interpretation of the data, owing to lack of purity of the mRNA and to an incomplete knowledge of the specificity of hybridisation. The results of such studies have been taken to support both germ-line (Premkumar et al, 1974) and somatic theories (Tonegawa et al, 1974).

1.2.6 Anti-NIP antibodies

Kreth & Williamson (1973) have attempted to estimate the minimum number of anti-NIP (4-hydroxy-3-iodo-5-nitro-phenacetyl-) antibodies that can be elicited in CBA/H mice. Cell-transfer at limiting dilution from mice primed with NIP coupled to bovine IgG into irradiated recipients permitted visualisation of individual clones by isoelectric focusing of the serum in polyacrylamide gels and overlaying with labelled antigen. The low frequency of identical isoelectric focusing patterns (assumed to reflect identical sequences) was interpreted as indicating a minimum number of 70 V_H and 70 V_L genes specifying anti-NIP. The repeat patterns are suggestive of inherited structural genes coding for those sequences. However the majority (332 out of 337) of patterns were unique and could have arisen by somatic mutation.

1.2.7 Rate of mutation

Somatic theories of diversity make certain predictions as to the nature and frequency of mutation in Ig structural genes. Random mutation followed by selection requires a minimum rate of point mutations giving rise to amino acid substitutions (10^{-6}, Cohn 1971; 10^{-7}, Jerne 1971). Hypermutation would result in a non-random pattern of mutations with more substitutions in the hypervariable region and recombination would also produce characteristic mutations. These predictions had not been tested by any direct observation of somatic mutation in Ig genes. Indeed until the results of chapter 6 were first reported (Secher et al, 1973), no spontaneous somatic mutation in the cells of a higher organism had been correlated with an altered amino acid sequence. Instead assumptions about the rate and nature of mutation in Ig genes were made largely by extrapolation from prokaryotic systems, although the arrangement of genes in higher

organisms is known to be much more complex (Britten & Kohne, 1967).

Burnet (1969) suggested that myeloma cells would serve as a model system for studying the somatic variation of normal antibody-producing cells, except that "it would be impossible in practice to detect the fraction of 1% which had undergone a significant change" (p.144). Chapters 2 and 5 of this dissertation describe the development of a suitable technique for screening large populations of myeloma cells in culture. Spontaneously arising variants may be isolated, propagated and further studied. Chapter 6 describes some structural studies of the variant Ig's produced by two such clones in an attempt to characterise the nature of the mutation.

1.2.8 Previous studies of somatic mutation

Most studies of somatic mutation have concerned a phenotypic property e.g. sheep fleece (Fraser & Short, 1958) or the loss of an enzymic acitivity (e.g. Breslow & Goldsby, 1969; Chu et al, 1969). In view of the complexities of control of eukaryotic gene expression, the loss of an enzymic function could be due to many other factors than a structural gene mutation and this may partly explain the wide variation in observed muatation rates (Coffino & Scharff, 1971). The difficulty of distinguishing structural gene mutations from control gene mutations can be partially overcome by detecting electrophoretic variants (Tobari & Kojima, 1972). However without chemical analysis of the variant product, it cannot be known that the alteration in charge is not due to, say, the absence of carbohydrate adding enzymes.

Scharff and his collaborators have studied somatic mutation in myeloma cells by screening for the cessation of H-chain secretion (Coffino et al, 1972). They found a rate of 1.1×10^{-3} /cell /generation for the conversion of H+L-chain to L-chain secretion in the cell lineMPC-11 (Coffino & Scharff, 1971). No H-chain was

detected inside or outside the cell. More recently mutagenesis has been used and some mutants producing, but not secreting, H-chains of altered size were detected (Preud'homme et al, 1973). These may be structural gene mutations but confirmation requires chemical analysis of the wild type and variant H-chains.

Finally it might be observed that as the techniques which promise to resolve the problem become more refined the differences between the theories is decreasing. If further sequence data for mouse κ chains were to suggest 100 V_κ genes, how many more would be required by a germ-line theory if one allows for multiple specificities?

CHAPTER 2

Materials and Methods

2.15 Peptide fingerprinting techniques

2.15.1 Electrophoresis

2.15.2 Chromatography

2.15.3 Location of peptides

2.16 Amino acid analysis

2.17 Dansyl-Edman degradation

2.18 Radioactivity measurements

CHAPTER 2

Materials and Methods

2.1 Materials

Tissue culture media constituents were obtained from Flow Laboratories Ltd., Irvine, Ayrshire, UK. Chemicals used in the Dansyl-Edman technique were "Sequenal" grade from Pierce Chemical Company, Rockford, Ill., USA or were redistilled according to Edman & Begg (1967). Agar ("Bacto-Agar") was from Difco Laboratories, Detroit, Mich., USA. L-[U-^{14}C] lysine monohydrochloride (336mCi/mmol), L-[U-^{14}C]-arginine monohydrochloride (324mCi/mmol), L-[^{35}S]cystine (195mCi/mmol) and iodo[2-^{14}C]acetate (10-25mCi/mmol) were purchased from the Radiochemical Centre, Amersham, Bucks., UK. Trypsin, pepsin and carboxypeptidase A were from the Worthington Biochemical Corp., Freehold, N.J., USA; neuraminidase (Type VI from Clostridium perfringens) and soybean trypsin inhibitor from Sigma London Chemical Company Ltd., Kingston-upon-Thames, Surrey, UK. Myeloma protein was prepared from serum kindly provided by Mr D.P. Wright. Sheep antiserum was a gift from Dr C.P. Milstein. Amphotericin B ("Fungizone") was from E.R. Squibb & Sons, London, UK). Ampholines were obtained from LKB Produkter AB, Bromma, Sweden. Acrylamide and N,N'-methylene bisacrylamide were recrystallised before use. Other materials, usually "Analar" grade, were used without further purification. "Kodirex Auto-Process" and "Blue Brand" X-ray films for radioautography were from Kodak Ltd., Hemel Hempstead, Herts., UK and were processed in a Kodak "X-omat" automatic processor.

2.2 Cell lines

The cell lines used all derive from the mouse plasmacytoma MOPC21, also known as P3 (Potter, 1972). The tumour was adapted to tissue culture

and cloned *in vitro* by Horibata & Harris (1970). The cultures used in this study were established from clone P3K 6.2.8.1., which was a gift from Dr K. Horibata. In this thesis the abbreviation P3K is used to refer to cells of this clone and also to the solid tumours which produce the same Ig (Svasti & Milstein, 1972a) and which are maintained by serial passage in Balb/c mice.

During the course of the work a number of clones were isolated from P3K cultures. These were classified as "wild-type" if they appeared to secrete the same Ig as the parent culture; "non-producers" if no Ig secretion was detectable; and "mutants" if they secreted an Ig of altered pI. IF-1 and IF-2 are two such mutant clones which were studied in detail (Chapter 6).

2.3 Solid tumours

The MOPC 21 plasmacytoma was maintained by serial transplantation into Balb/c mice. Transfer of tumours was achieved as follows. Excised solid tumours were chopped up in 3-5 ml Eagle's medium (Eagle, 1959) and 0.2 ml of the supernatant fluid (containing about 10^7 cells) was injected subcutaneously into a normal Balb/c mouse. After about 11 days, when a solid tumour had developed around the point of injection, the transfer operation was repeated.

IF-1 and IF-2 were also maintained as solid tumour lines. To establish the IF-1 line, 10^7 tissue culture cells, in 0.2 ml growth medium (see below) were injected into each of 20 Balb/c mice. A small proportion of mice developed visible tumours which were transferred into healthy mice as described above. The tumour line appeared to "adapt" to growth in mice in that at each serial transfer a higher proportion of inoculated animals developed tumours until, after four or five transfers, the efficiency reached 100%.

IF-2 cells, when similarly injected into Balb/c mice, produced no detectable tumours. In an attempt to suppress the mechanism responsible

for this failure, mice were subjected to whole-body irradiation from a ^{60}Co source, shortly before injection of cells. Out of 14 mice subjected to 400 rad, one developed a solid tumour. This was transferred, in the normal way, to 10 mice which had been irradiated with 350 rad. Two of these animals developed tumours. Subsequent transfer to 10 mice irradiated with 300 rad resulted in tumours in 9 animals. Further transfers into non-irradiated mice were successful and eventually reached 100% efficiency. Three lines of IF-2 tumours have been established using this procedure.

The sera of mice carrying IF-1 and IF-2 tumours contained high levels of the corresponding myeloma protein as judged by cellulose acetate electrophoresis. The yield of protein (2-5mg/ml serum) purified from these sera was lower however than from sera obtained from mice bearing wild type P3K tumours (20-25mg/ml, Svasti, 1972).

2.4 Culture methods

2.4.1 Media

Cells in continuous culture were grown in 1ℓ or 3ℓ spinner bottles in Dulbecco's modified Eagle's medium (Smith et al, 1960), supplemented with 10% (v/v) heated horse serum ("growth medium"). The cultures were maintained under an atmosphere of $N_2+O_2+CO_2$ (83:7:10, BOC Ltd., Brentford, Middx., UK). Medium for cloning (Medium C) contained 20% (v/v) heated horse serum but was otherwise similar to growth medium. For incorporation of radioactive amino acid(s) media lacking the appropriate amino acid(s) were prepared (e.g. Medium C (minus lysine)).

2.4.2 Cloning

Cells were cloned essentially according to the method outlined by Paul (1970). 15ml agar (0.5% (w/v) in Medium C) was solidified in a

90 mm plastic Petri dish and overlaid with cells in 2 ml soft agar (0.25% (w/v) in Medium C). Petri dishes were incubated for at least 11 days at 37°C under an atmosphere of air + CO_2 (9:1) in a humidified incubator. By this time colonies were visible with the naked eye and were removed, with a sterile Pasteur pipette for further study (see below) or transferred for continued growth into 2 ml cups (24 cup plate, Linbro Chemical Co., New Haven, Conn., USA) where they were maintained in Medium C at a density of 10^4-$5x10^5$ per ml. Cloning efficiency was about 1%.

2.4.3 Detection of non-producer clones

Clones were grown as above, but after 3 days 1 ml of Medium C, lacking horse serum but containing 0.25% (w/v) agar and 20% (v/v) sheep antiserum raised against purified P3K IgG, was solidified over the colonies. After several days further incubation, colonies producing Ig became surrounded by precipitate (Coffino & Scharff, 1971).

2.5 Incorporation of ^{14}C-lysine into secreted Ig

2.5.1 Screening Methods

Two methods of incorporating labelled amino acid for screening of cells were used. The first destroyed the clones studied, but was preferred intially as it was more rapid.

2.5.1.1 Incorporation in agar Whole agar plates containing 30-60 colonies larger than 1 mm in diameter were overlaid with 1.4 ml Medium C (minus lysine) containing 10 µCi ^{14}C-lysine and 0.36% (w/v) agar. The plates were incubated at 37°C for a further 24 hr. Each colony in a plug of surrounding agar was transferred with a Pasteur pipette onto a polyacrylamide gel slab for isoelectric focusing.

2.5.1.2 Incorporation on dialysis membrane (Fig.5.4) - 16.4 ml Medium C (minus lysine) containing 7.5 µCi ^{14}C-lysine and 0.5% (w/v) agar was solidified in a plastic dish (14.5 cm x 10 cm). Strips of dialysis tubing (14 cm x 1 cm), Visking $^{20}/_{32}$", Scientific Instrument Centre Ltd., London, UK) were laid on the agar. Colonies picked from cloning plates with sterile Pasteur pipettes were placed on the strips. Each pipette was then used to inoculate 0.2 ml Medium C containing 2.5 U/ml Amphotericin B in a separate cup (Linbro 96 cup plate, Bio-cult Laboratories Ltd., Paisley, Renfrewshire, UK). The incorporation dish was incubated for 24 hr. at 37°C in the humidified incubator and for 1.5 hr. at 37°C in a dry incubator. The strips were then placed, colony side down, on the origin of a polyacrylamide slab for isoelectric focusing.

2.5.2 Incorporation by cells in suspension culture

2.5.2.1 Analytical 0.1×10^6-1×10^6 cells were suspended in 1 ml Medium C (minus lysine) in sealed plastic tubes and incubated at 37°C for 15 min. After centrifugation, the cells were resuspended in 0.1 ml medium C (minus lysine) containing 0.42 µCi ^{14}C-lysine. The tubes were gassed with N_2+O_2+CO_2 (83:7:10) and incubated at 37°C, usually for 24 hr. The cells were removed by centrifugation and the supernatant stored at -20°C.

2.5.2.2 Preparative To prepare radioactively labelled protein for fingerprinting studies, 10^6-10^7 cells were washed twice in growth medium lacking the appropriate amino acid(s) and resuspended in 10 ml of the same medium and transferred to a 250 ml tissue culture flask. After addition of 300 µCi ^{35}S-cystine or 50 µCi ^{14}C-lysine and 50 µCi ^{14}C-arginine the cells were incubated for 4 days at 37°C. At 24 hr., 48 hr. and 72 hr. 1 ml of complete growth medium containing 10^5-10^6 actively growing cells was added. At 96 hr. 0.25 ml mouse serum

2.5.1.2 Incorporation on dialysis membrane (Fig.5.4) - 16.4 ml Medium C (minus lysine) containing 7.5 µCi ^{14}C-lysine and 0.5% (w/v) agar was solidified in a plastic dish (14.5 cm x 10 cm). Strips of dialysis tubing (14 cm x 1 cm), Visking $^{20}/_{32}$", Scientific Instrument Centre Ltd., London, UK) were laid on the agar. Colonies picked from cloning plates with sterile Pasteur pipettes were placed on the strips. Each pipette was then used to inoculate 0.2 ml Medium C containing 2.5 U/ml Amphotericin B in a separate cup (Linbro 96 cup plate, Bio-cult Laboratories Ltd., Paisley, Renfrewshire, UK). The incorporation dish was incubated for 24 hr. at 37°C in the humidified incubator and for 1.5 hr. at 37°C in a dry incubator. The strips were then placed, colony side down, on the origin of a polyacrylamide slab for isoelectric focusing.

2.5.2 Incorporation by cells in suspension culture

2.5.2.1 Analytical 0.1×10^6-1×10^6 cells were suspended in 1 ml Medium C (minus lysine) in sealed plastic tubes and incubated at 37°C for 15 min. After centrifugation, the cells were resuspended in 0.1 ml medium C (minus lysine) containing 0.42 µCi ^{14}C-lysine. The tubes were gassed with N_2+O_2+CO_2 (83:7:10) and incubated at 37°C, usually for 24 hr. The cells were removed by centrifugation and the supernatant stored at -20°C.

2.5.2.2 Preparative To prepare radioactively labelled protein for fingerprinting studies, 10^6-10^7 cells were washed twice in growth medium lacking the appropriate amino acid(s) and resuspended in 10 ml of the same medium and transferred to a 250 ml tissue culture flask. After addition of 300 µCi ^{35}S-cystine or 50 µCi ^{14}C-lysine and 50 µCi ^{14}C-arginine the cells were incubated for 4 days at 37°C. At 24 hr., 48 hr. and 72 hr. 1 ml of complete growth medium containing 10^5-10^6 actively growing cells was added. At 96 hr. 0.25 ml mouse serum

containing myeloma protein was added, the cells removed by centrifugation and the labelled protein prepared from the extracellular fluid by ion-exchange chromatography on DEAE-cellulose as described below. In this way up to 18 µCi (18% of input label) of ^{14}C-labelled protein or 7 µCi (2.5% of input label) of ^{35}S-labelled protein was prepared in a single experiment.

2.6 Incorporation of ^{14}C-lysine into intracellular Ig

About $5x10^6$ cells were washed in Medium C (minus lysine), resuspended in 5 ml of the same buffer and preincubated at 37°C for 15 min. The cells were removed by centrifugation and resuspended in 0.25 ml Medium C (minus lysine) containing 10 µCi of ^{14}C-lysine. After a further 15 min. at 37°C the cells were harvested and lysed by addition of sodium deoxycholate (Awdeh et al, 1970) or "Nonidet P40" to 1% (v/v) final concentration.

2.7 Isoelectric focusing

Isoelectric focusing was performed in polyacrylamide slabs, essentially according to Awdeh et al. (1968). The composition of monomer solution for one plate was:

	ml
Monomer solution (19% (w/v) acrylamide, 0.4% (w/v) N,N'-methylene bisacrylamide)	10
6M urea	20
0.004% (w/v) riboflavin	2
Ampholine (40% (w/v) solution as supplied)	1.3
1.25% (v/v) N,N,N',N'-tetramethylene diamine (TEMED)	0.67

Ampholine of pH range 6-8 was normally used. In later experiments it was replaced with a mixture of pH 6-8, pH 5-7 Ampholines (13:9(v/v)). The broad range pH 3-10 Ampholine, used at 0.67 ml/25 ml gel solution, was also used occasionally.

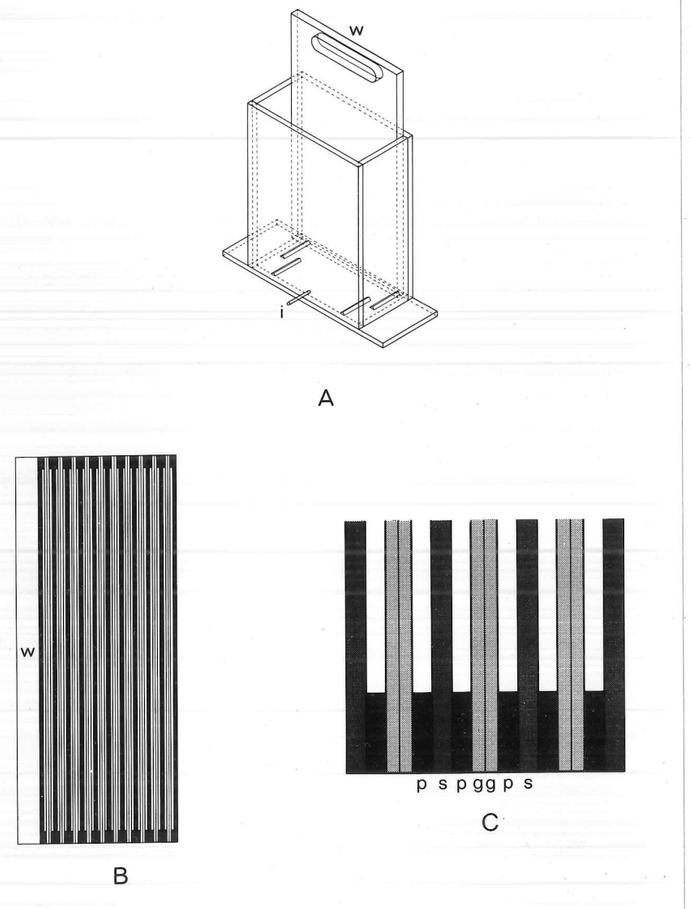


Fig. 2.1 Apparatus for preparation of polyacrylamide gels for isoelectric focusing. A) Sketch of perspex box, with "sandwich" removed, showing wedge (w) used to compress the sandwich, and the inlet (i) for the gel solution. B) Cross-section of sandwich with wedge. The blank areas within the sandwich represent the spaces in which the 20 gels polymerise. C) An enlarged detail of B showing arrangement of perspex spacer strips (p), siliconised glass plates (s) and glass plates coated with gelatin (g).

The solution was degassed before the addition of the TEMED and then poured between two glass plates, one of which had been coated with dimethyldichlorosilane ("Repelcote", Hopkin & Williams Ltd., Romford, Essex, UK), the other with gelatin (glass prepared for coating, 16.5 cm x 21.6 cm, Ilford Ltd., Ilford, Essex, UK). The glass plates were either sealed along three edges with a length of 0.8 mm polyethylene tubing held in place with "Foldback" paper clips or were held apart by perspex strips (215 mm x 5 mm) and stacked in a multilayer "sandwich" containing space for 20 gels (Fig.2.1).

This sandwich was placed in a reservoir (Fig. 2.1) into which 75 ml water, 550 ml gel solution and 50 ml 50% sucrose were successively poured through an entry in the base of the reservoir. The gels were polymerised by overnight exposure to a fluorescent light tube (40W) at a distance of 10-20 cm . After polymerisation the complete sandwich of gels and plates was removed from the reservoir. Single gels were removed as required; the unused gels were stored in an airtight polyethylene bag at $4^{o}C$ in the dark for up to six months.

Samples were applied about 10 cm from the anode. For the analysis of material secreted during an incorporation by cells in suspension, 10-50 µl of extracellular incorporation medium was applied to a strip of Whatman No.1 paper (about 1 cm x 0.5 cm) which was then placed on the gel at the origin. Up to 30 samples per gel were applied in this way. The gels were run for 18 hr. - 24 hr. (250V for 1 hr., 400V (22.5V/cm) for the remaining time) at $4^{o}C$ in an apparatus specially constructed to take 6 plates (Fig. 2.2). The inverted plates were supported on carbon electrodes and uniform electrical contact ensured by placing strips of Whatman No.1 paper (18 cm x 1 cm) presoaked in 5% H_3PO_4 (anode) or 5% ethanolamine (cathode). Staining and destaining using bromophenol blue were carried out as described by Awdeh (1969). The destained gels were air dried at room temperature and radioautographed for 0.5-3 days.

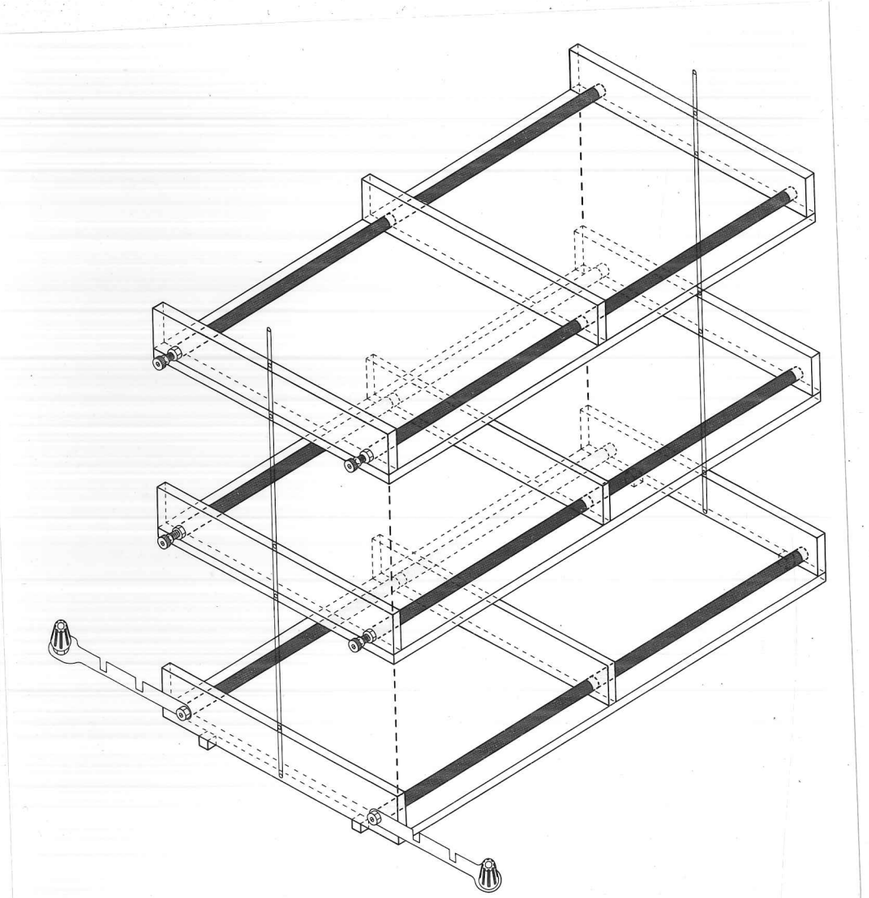


Fig. 2.2 Exploded diagram of apparatus for running 6 isoelectric focusing slab gels simultaneously. Constructed from perspex with carbon rod electrodes (stippled) and brass electrical terminals. Approximate dimensions when assembled 40cm x 22cm x 12cm. Whole apparatus placed in humid chamber for running.

2.8 mRNA preparation

mRNA was prepared in a partially purified state by a simplified version of the method of Brownlee et al (1973). All operations were carried out at 4°C unless otherwise stated. 2.6 ℓ of IF-1 cells (3×10^5 cells/ml) were harvested from a spinner culture by centrifugation at room temperature, washed in ice-cold 0.01M Tris-HCl pH 7.5, 0.01 M KCl, 0.001 $MgCl_2$ and resuspended in 25 ml of the same (low-salt) buffer, giving a total volume of 2.5 ml. This suspension was subjected to 10 strokes in a tight-fitting Dounce homogeniser to disrupt the cells. The homogenate was diluted with 2 ml 1.5M sucrose, 0.11M Tris-HCl pH 7.5, 0.05M KCl, 0.001 $MgCl_2$ and centrifuged at 600g for 5 min. The pellet (nuclear fraction) was washed with 3 ml 0.6M sucrose, 0.05M Tris-HCl pH 7.5, 0.025M KCl, 0.001 $MgCl_2$ and recentrifuged (600g, 5 min.) as was the supernatant from the previous centrifugation. The final supernatants of both tubes were pooled and the microsomal fraction prepared by a further centrifugation at 30,000g for 7 min. The microsomal pellet was washed in 5 ml 0.05M Tris HCl pH 7.5, 0.5M KCl, 0.002M $MgCl_2$ and recentrifuged under the same conditions. The final microsomal pellet was hand-homogenised in 3 ml 3% (w/v) SDS, 0.01M Tris-HCl pH 7.5, 0.1M NaCl, 0.005M EDTA and incubated at 37°C for 15 min. 1.5 ml of the clear solution was applied to each of two 35 ml sucrose gradients 15-30% (w/v) ~~EDTA~~ sucrose in 0.5% (w/v) SDS, 0.01M Tris-HCl pH 7.5, 0.1M NaCl, 0.005 M EDTA and centrifuged at 90,000g for 19 hr. at 27°C (26,000 rpm in a Beckman SW27 rotor). The gradients were pumped out and monitored for absorbance at 254 nm; 2.8 ml fractions were collected. RNA was prepared from fractions in the 12S-20S region by addition of 7 ml ethanol and storage at -20°C overnight. The RNA was recovered by centrifugation and dried in vacuo. Traces of SDS were removed by reprecipitating the RNA from 2% (w/v) sodium acetate pH 7.0 with 2.5 volumes of ethanol as described above.

2.9 In vitro assay of mRNA

RNA fractions were assayed for mRNA activity by their ability to direct the synthesis of proteins in an enriched rabbit reticulocyte lysate (Stavnezer & Huang, 1971). The reticulocyte lysate, prepared essentially according to Lockhard & Lingrel (1969) was a gift from Dr T. M. Harrison. The reaction mixture for protein synthesis was prepared as follows:

	µl
lysate	285
haemin (600 µM)	15
1.4M KCl, 0.02M $MgCl_2$	25
0.1M ATP, 0.01M GTP, 1.5M creatine phosphate	5
amino acids (minus methionine) (2mM each)	5
^{35}S-methionine	33
creatine phosphokinase	a few crystals
water	to 400

Each protein synthesis reaction consisted of 40 µl of the above mixture plus 10 µl RNA solution in water and was incubated at 30°C for 1 hr. The reaction was stopped by addition of 2 µl pancreatic RNase (1 mg/ml) in 90 mM EDTA, pH 7.5 and further incubation at 37°C for 15 min. The mixture was applied directly to an SDS-polyacrylamide gel slab for analysis of the radioactive proteins.

2.10 SDS gel electrophoresis

The procedure used has been described in detail by Laemmli & Faure (1973). The slab gels used offer considerable advantages over cylindrical gels especially when different samples are to be compared. The apparatus is relatively simple to construct and is now commercially available (Raven Scientific Instruments Ltd., Haverhill, Suffolk, UK). Electrophoresis in 12% polyacrylamide was usually overnight at 5-10 mA. Gels were then stained with Coumassie brilliant blue. For the analysis

of labelled protein the slabs were dried onto a sheet of chromatography paper in an evacuated polythene bag and then radioautographed.

2.11 Purification of myeloma protein

Myeloma protein was prepared from mouse serum by precipitation with ammonium sulphate, to 40% saturation, followed by ion-exchange chromatography on a column of DEAE-cellulose (20cm x 2cm; DE52, Whatman Biochemicals Ltd., Maidstone, Kent, UK) equilibrated with 0.01M sodium phosphate, pH 7.0 (Svasti & Milstein, 1972a). IF-1 IgG passed unretarded through the column; P3K IgG was eluted by application of a linear sodium phosphate gradient (0.01M - 0.2M, pH 7.0, total volume = 400 ml).

The purification of labelled myeloma protein from cell culture medium was similar, but ammonium sulphate precipitation was omitted and the ion exchange chromatography preceded by desalting the medium on Sephadex G-25 (coarse, 80 cm x 2.5 cm) equilibrated with 0.01M sodium phosphate pH 7.0. The protein peak from this column was applied to a 6 cm x 2 cm column of DE52 and the myeloma protein eluted as above.

Purity of myeloma preparations was determined by electrophoresis on cellulose acetate strips ("Microzone", Beckman Instruments Inc., Palo Alto, Calif., USA). Electrophoresis was for 1 hr. at 20V/cm in 0.05M veronal buffer pH 8.6.

2.12 Reduction of disulphide bridges

Selective reduction of interchain disulphide bridges with dithiothreitol and carboxymethylation with iodoacetic acid were as described by Frangione et al (1969) except where modified by Svasti & Milstein (1972a). H- and L-chains were separated by gel filtration on Sephadex G-100 (80 cm x 3 cm) in 5% (v/v) formic acid. Total reduction and alkylation was performed in 6.6M guanidine as described by Pink et al (1970).

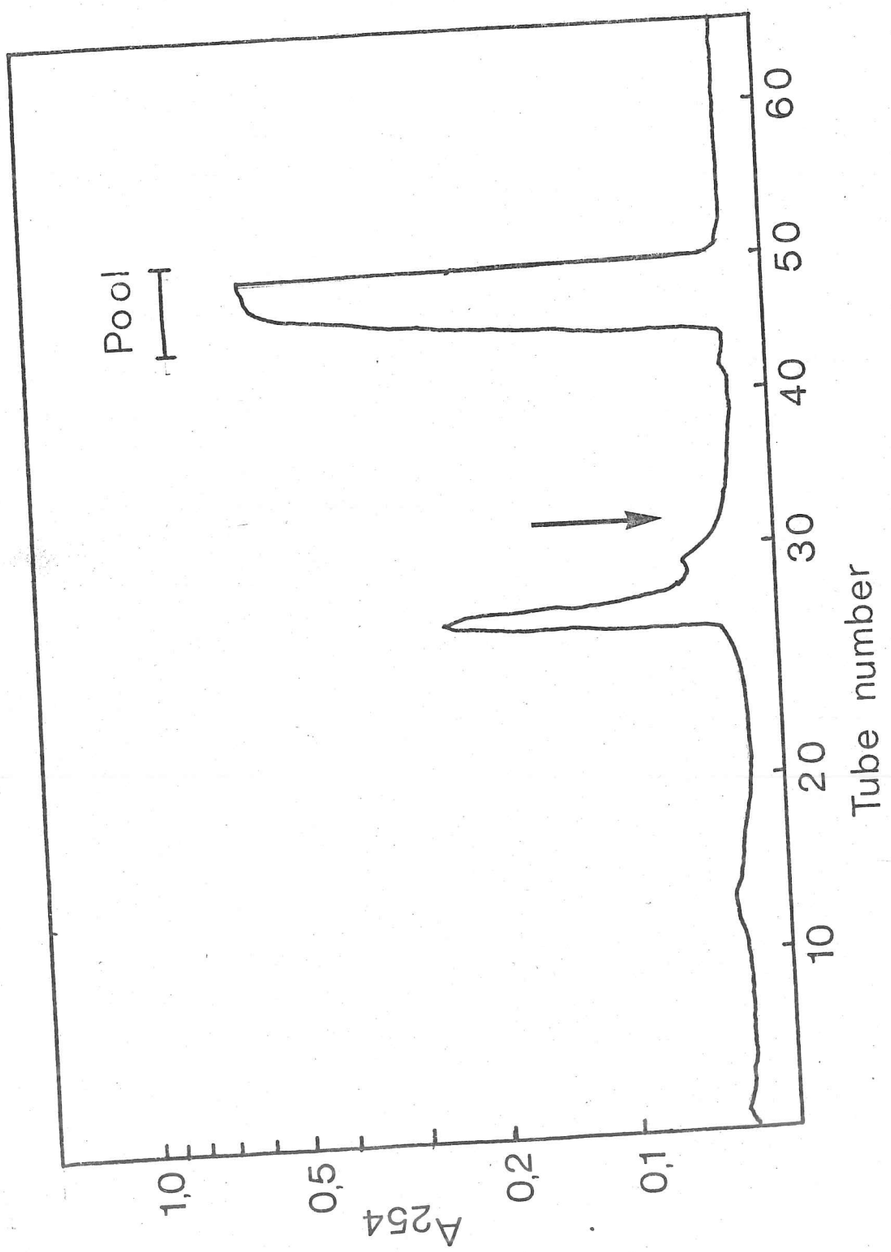


Fig. 2.3 Purification of trypsin by affinity chromatography on soybean trypsin inhibitor. The column (12cm x 2.5 cm) was loaded with 85mg trypsin (Worthington, Batch TRL6261) dissolved in 8ml tris-HCl pH 7.1, 0.02M $CaCl_2$ and eluted with the same buffer. 150 drop fractions were collected and after the elution of the first peak (unretarded material) the eluting solution was changed to 0.003M HCl, 0.05M $CaCl_2$ (arrow). The second peak, containing 95% of the initial tryptic activity was pooled as shown.

2.13 Performic acid oxidation

Performic acid was prepared by mixing 1 volume of 30% (v/v) H_2O_2 with 19 volumes of formic acid. After 2 hr. at room temperature, 2 volumes of reagent was added to 1 volume of formic acid in which the sample to be oxidised had been dissolved. After 45 min. at room temperature the reaction mixture was diluted with water (about 20 volumes) and lyophilised.

2.14 Enzymatic digestion

2.14.1 Trypsin

Trypsin was purified before use by affinity chromatography on soybean trypsin inhibitor linked to Sepharose 4B (a gift from Dr J. Sperling). The column (12 cm x 2.5 cm) was equilibrated with 0.1M Tris-HCl pH 7.1, 0.02M $CaCl_2$ and the trypsin (85 mg) loaded in 8 ml of the same buffer. The column was eluted with the same buffer until the eluent was free of protein (detected by absorbance at 254nm). The column was then eluted with 0.003M HCl, 0.05M $CaCl_2$ (Fig. 2.3). The protein peak eluted by the acid contained all the tryptic acitivity and was divided into 2.5 ml aliquots (1 mg/ml protein concentration and stored at $-20^{\circ}C$. 95% of the tryptic activity was recovered with a purification factor of 1.33.

Tryptic activity was assayed by the method of Schwert & Takenaka (1955).

Tryptic digests of separated H- and L-chains were carried out at $37^{\circ}C$ and 10 mg/ml substrate concentration for 4 hr.; enzyme:substrate ratio was 1:100. Digests were terminated by the addition of a weight of soybean trypsin inhibitor equal to the weight of trypsin used.

2.14.2 Pepsin

For fingerprinting, peptic digestion of whole IgG was

carried out for 16 hr. at 37°C in 5% formic acid and with an enzyme: substrate ratio of 1:40.

To produce large fragments by limited proteolysis, whole IgG was digested in 0.1M sodium acetate pH 4.0 for 16 hr. at 37°C. The products of this reaction were analysed by SDS gel electrophoresis.

2.14.3 Carboxypeptidase A and B

C-terminal analysis of whole IgG was performed using carboxypeptidases A and B sequentially. Carboxypeptidase A was activated as described by Ambler (1972). The release of free amino acids as a function of time was measured by automatic amino acid analysis.

2.14.4 Neuraminidase

Neuraminidase digestion, to specifically remove sialic acid (Cassidy et al, 1965) was performed on ^{14}C-lysine labelled protein prior to isoelectric focussing for 4 hr. with an enzyme concentration of 10 µg/ml in growth medium (pH about 7).

2.15 Peptide Fingerprinting Techniques

2.15.1 Electrophoresis

The following buffers were used for high voltage paper electrophoresis, performed essentially as described by Ambler (1963) or Milstein (1966b) using Whatman No.1 paper (for loading up to 0.5 mg/cm) or 3MM paper (up to 1mg/cm).

pH 2.0	8% (v/v) acetic acid, 2% (v/v) formic acid
pH 3.5	5% (v/v) " " , 0.5% (v/v) pyridine
pH 6.5	0.3% (v/v) " " , 10% (v/v)
pH 8.9	1% (w/v) ammonium bicarbonate

As external markers a mixture (1-5 µl) of the amino acids lysine, arginine, histidine, glycine, valine, ε-DNP-lysine, glutamic acid,

aspartic acid, taurine and cysteic acid, the dipeptide glycyl-alanine (all at 5mM final concentration) and the dye Xylene Cyanol FF (amount judged visually) was applied at the origin near the sample.

Mobilities, expressed as fractions of the distance between ε-DNP-lysine and aspartic acid after electrophoresis at pH 6.5 were used to assign amide groups (Offord, 1966).

2.15.2 Chromatography

Descending paper chromatography was carried out using the solvent system butan-1-ol, acetic acid, water, pyridine (15:3:12:10 by volume) (BAWP) (Waley & Watson, 1953). "Rf" values were measured relative to valine.

2.15.3 Location of peptides

Radioactive peptides were detected by radioautography. Non-radioactive peptides were detected by their fluorescence under ultra-violet light (for tryptophan containing peptides) or by staining with ninhydrin. 0.003% (Anfinsen et al, 1959) - 0.01% (w/v) ninhydrin in acetone was used when peptides were to be recovered for amino acid analysis. The cadmium/ninhydrin reagent of Heilmann et al (1957) was used to detect other peptides.

2.16 Amino acid analysis

HCl for total acid hydrolysis of peptides and proteins was prepared by diluting 1 volume of Aristar grade HCl (34-36% solution, B.D.H. Chemicals Ltd., Poole, Dorset, UK) with 1 volume of distilled water and adding phenol (final concentration 1 mg/ml) to reduce destruction of tyrosine (Sanger & Thompson, 1963).

A Beckman 120 amino acid analyser, modified by the addition of a

Locarte (London, UK) automatic loading, single-column system and high sensitivity flow cells,was used for all amino acid analyses.

2.17 Dansyl-Edman degradation

Proteins were subjected to automatic Edman degradation (Edman & Begg, 1967) in a Beckman 890B Sequencer. Details of the program and identification procedures are given in Chapter 3 .

The dansyl-Edman technique as described by Gray (1972) was used to determine peptide sequences. The peptide was dissolved in 100 µl water in a 3 ml screw cap culture tube fitted with a P.T.F.E. seal. 100 µl phenyl isothiocyanate (5% (v/v) in pyridine) was added and the tube flushed with nitrogen. After 75 min. at 45°C the solvents and volatile reagents were removed under vacuum over P_2O_5 and NaOH at 60°C and 150 µl of trifluoracetic acid added. The tube was incubated for 30 min. at 45°C and the volatile acid removed under vacuum over NaOH at 60°C. 150 µl water was added to dissolve the peptide and extracted three times with 1.5 ml butyl acetate. An aliquot of the remaining peptide (1-5 nmol), now one residue shorter was removed for reaction with Dns-Cl and the remainder dried under vacuum.

The sample of peptide removed for amino terminal determination was dried in a pointed Pyrex tube (4 cm x 0.4 cm, Wesley Coe Ltd., Cambridge, UK) and redissolved in 5 µl 0.2M $NaHCO_3$. 5µl Dns-Cl (2.5 mg/ml in acetone) was added, the tube covered, and incubated at 45°C for 45 min. After drying, the derivatised peptide was hydrolysed with 10 µl 6N HCl at 105°C for 4-16 hr., dissolved in 1 µl 50% (v/v) pyridine and applied equally to both sides of a polyamide thin-layer sheet (B.D.H. Chemicals Ltd.). A marker containing the Dns-derivatives of Pro, Ile, Phe, Gly, Ser, Glu and Arg was applied to one side only to act as an internal standard.

The following solvent systems were used to separate the Dns-amino acids:

Solvent I: formic acid, water (3:200) (Woods & Wang, 1967)

Solvent II: benzene, acetic acid (9:1) (Woods & Wang, 1967)

Solvent III: ethyl acetate, acetic acid, methanol (20:1:1) (Crowshaw et al, 1967)

Solvent IV: "pH 4.4 buffer":ethanol (7:3)

("pH 4.4 buffer" is 0.8% (v/v) pyridine, 1.6% (v/v) acetic acid (1:1) (Gray & Hartley, 1963))

Solvent V: 0.05M Na_3PO_4, ethanol (3:1)

Solvents IV and V are attributed to Dr S. Magnusson (Svasti, 1972).

The plates were developed with solvent I for about 20 min. and then, after drying, with solvent II in a perpendicular direction. These two solvents resolved all Dns-amino acids except the derivatives of Asp/Glu, Thr/Ser and the basic amino acids. Solvent III was used to separate Dns-Thr from Dns-Ser and Dns-Glu from Dns-Asp. Solvent IV separated α-Dns-His from ε-Dns-Lys/α-Dns-Arg and Solvent V resolved ε-Dns-Lys from α-Dns-Arg.

Plates were reused after overnight washing in ethyl acetate, propan-1-ol, ammonia (35% (w/v), water (10:6:1:3 by volume).

2.18 Radioactivity measurements

Radioactive samples were counted by liquid scintillation counting in a Nuclear Chicago Unilux I machine. The scintillant used was that of Bray (1960):

naphthalene	60g
PPO (2,5-diphenyloxazole)	4g
POPOP (1,4-di-2-(5 phenyloxazolyl)-benzene	0.2g
methanol	100 ml
ethylene glycol	20 ml
p-dioxan	to 1 ℓ

CHAPTER 3

The Use of the Sequenator

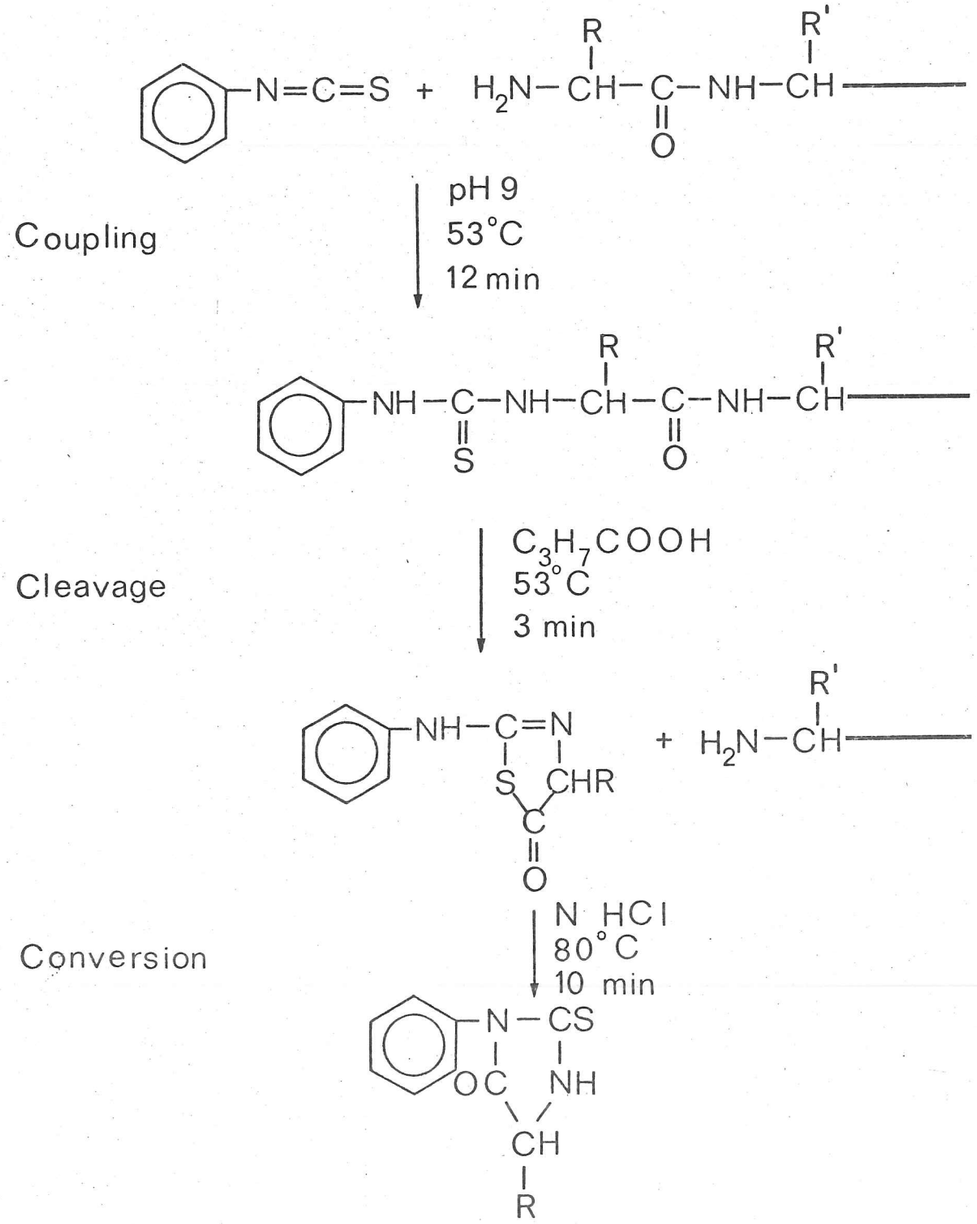


Fig. 3.1 Reaction scheme for Edman degradation as performed with the sequenator and conversion of thiazolinone to PTH-amino acid derivative. R and R' represent amino acid side chains.

CHAPTER 3

The Use of the Sequenator

3.1 Introduction

The development of an instrument for the automatic determination of amino acid sequences has undoubtedly been one of the major recent advances in protein chemistry (Edman & Begg, 1967). Since the introduction of commercial instruments in 1970 the technique has become available in more than 150 laboratories throughout the world. This chapter briefly describes the instrument, its capabilities and recent developments to the strategy of sequencing arising from the use of the sequenator.

3.2 Sequenator design

The design of the original sequenator is described in detail by Edman & Begg (1967). The machine automates the manual isothiocyanate degradation introduced by Edman (1950) and which has proved so successful in the analysis of small peptides (Fig. 3.1). In the sequenator the reactions take place inside a spinning glass cup in a thermostatically controlled gas-tight chamber which may be evacuated or filled with nitrogen. The entry and withdrawal of reagents, solvents and nitrogen and the application of vacuum are controlled by a punched-tape electronic programming unit. A simplified flow chart of the machine is shown in Fig. 3.2. The design has been copied, with only minor modifications by a number of instrument companies (see Science 174A, 133 (1971)) and by individual laboratories (e.g. Lynn & Bennet, 1972; Waterfield et al, 1970; Wittmann-Liebold, 1973b). An instrument manufactured by the Illinois Tool Co. has been used with considerable success by Smithies and his collaborators (Smithies et al, 1971) and a few machines have

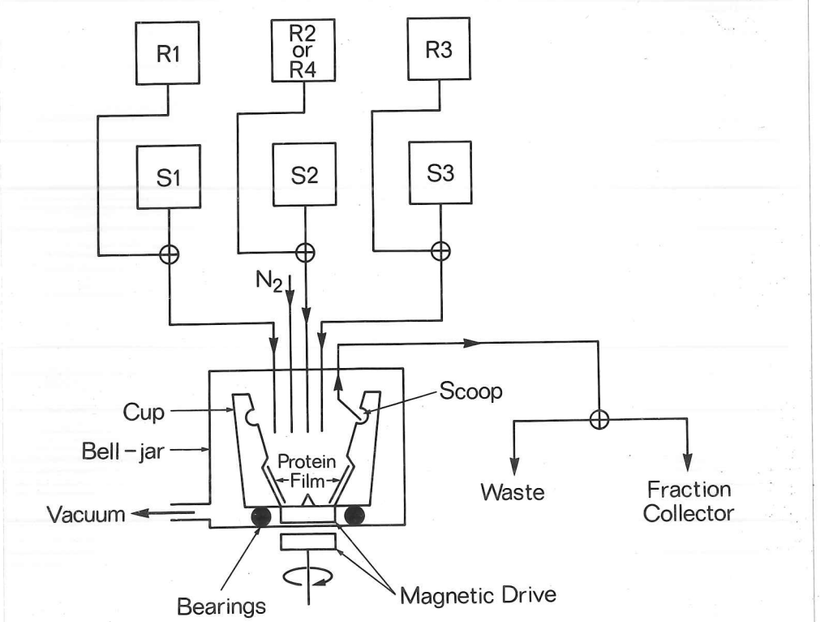


Fig. 3.2 Flow chart of sequenator (simplified). R1, R2, R3 represent reagent reservoirs (see text). R4 is the volatile buffer dimethyl-allylamine. S1, S2, S3 are solvent reservoirs. The nitrogen system which maintains all solutions oxygen-free and pressurises the reservoirs for delivery is not shown.

been sold by Socosi (France) and Jeol (Japan), but over 90% of the machines currently in use are produced by Beckman. The machine used in this laboratory is a Beckman "Sequencer 890B". Details of the machine's design are described elsewhere and need not be repeated here (Niall, 1973).

3.3 Modifications to instrument

To improve the drying characteristics of the protein films inside the glass cup, a "nitrogen bleed" valve was added. The programme was amended to activate this bleed allowing a fine stream of nitrogen into the cup during high vacuum steps. The valve was adjusted so that the vacuum under these conditions was about 400 mT. Certain other modifications were added by Beckman to incorporate the major developments available on a later model, 890C. The principal alteration was the replacement of the spinning cup with one possessing a "lip" half way up the side. This should improve the efficiency of the buffer layer and hence the buffering capacity with respect to evaporation (Fig. 3.3).

3.4 Chemicals

"Sequencer grade" reagents and solvents (Beckman) were used throughout the course of this work as follows:

R1 phenylisothiocyanate (5% (v/v) in heptane).

R2 M "Quadrol" [N,N,N',N'-tetrakis-(2-hydroxypropyl) ethylendiamine] buffered to pH 9.0 with trifluoroacetic acid in n-propanol/water (3:4).

R3 heptafluorobutyric acid

S1 benzene

S2 ethyl acetate

S3 1-chlorobutane

Acetic acid (B.D.H. Chemicals Ltd., Poole, UK "Aristar") was added to the S2 (1ml/ℓ) to improve the extraction of R2 from the protein film

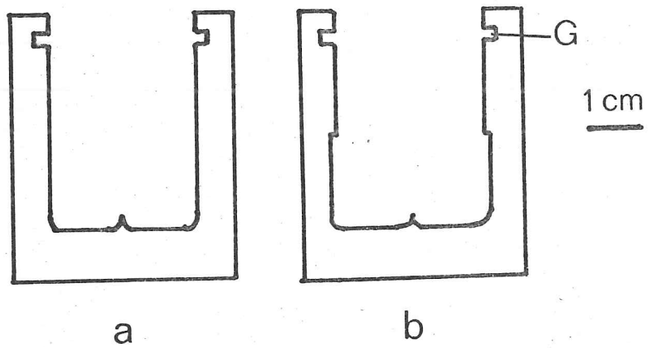


Fig. 3.3 Sequenator spinning glass cup in cross-section. a) Original design b) modified design with "undercut" cup (see text). The scale of the drawing is indicated by the bar which represents 1cm. G is groove from which solvents are removed. The scoop, plug, cup mounting and housing are not shown.

(Edman & Begg, 1967). Butane dithiol (Ralph Emmanuel, Wembley, UK) was added to S3 (50 µl/ℓ) to protect the extracted derivatives from destructive oxidation.

N HCl for conversion (B.D.H. Chemicals Ltd., Poole, UK, Aristar) contained 50 µl/ℓ ethane thiol (Pierce Chemicals, Rockford, Ill., USA, Hermodson et al, 1972). Sequencer grade ethyl acetate (Beckman Instrument Co. Inc., Palo Alto, USA) was also used for extraction of PTH-amino acids and as the GLC solvent.

3.5 Application of Sample

200-500 nmoles protein was usually used for a single run. Degradations on lesser amounts proceeded well, but identification using TLC and amino acid analysis became less informative. For degradations at the radiochemical level it was necessary to add an unlabelled "carrier" protein (e.g. 5mg sperm-whale apomyoglobin) as the low amounts of protein are otherwise extracted from the cup in the organic solvents. A similar finding has recently been reported by others (McKean et al, 1974; Niall et al, 1974). For rat Ig L-chains 20 mg was the upper limit - with larger amounts the protein flaked off the cup during drying steps.

For application the sample was dissolved in about 0.5 mls of a volatile solvent and dried by application of vacuum to the spinning cup. Water or 0.1M ammonia solution were generally preferred, but formic acid was used occasionally when the former failed to dissolve the sample. Solution of the sample, however, for application or during the degradation was not essential for efficient reaction. In cases where the sample was applied as a suspension, or where the R2 or R3 failed to dissolve the film, the success of the run appeared unaffected.

In some early runs a problem of detachment of the protein film during drying stages was encountered. This was later avoided by routine pretreatment of the protein film by solution in heptafluorobutyric acid and washing with chlorobutane according to the last part of the

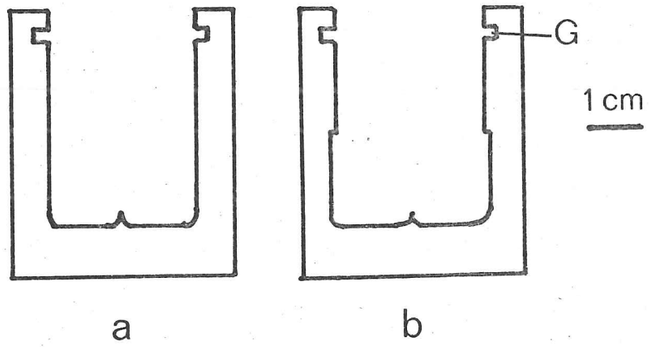


Fig. 3.3 Sequenator spinning glass cup in cross-section. a) Original design b) modified design with "undercut" cup (see text). The scale of the drawing is indicated by the bar which represents 1cm. G is groove from which solvents are removed. The scoop, plug, cup mounting and housing are not shown.

(Edman & Begg, 1967). Butane dithiol (Ralph Emmanuel, Wembley, UK) was added to S3 (50 µl/ℓ) to protect the extracted derivatives from destructive oxidation.

N HCl for conversion (B.D.H. Chemicals Ltd., Poole, UK, Aristar) contained 50 µl/ℓ ethane thiol (Pierce Chemicals, Rockford, Ill., USA, Hermodson et al, 1972). Sequencer grade ethyl acetate (Beckman Instrument Co. Inc., Palo Alto, USA) was also used for extraction of PTH-amino acids and as the GLC solvent.

3.5 Application of Sample

200-500 nmoles protein was usually used for a single run. Degradations on lesser amounts proceeded well, but identification using TLC and amino acid analysis became less informative. For degradations at the radiochemical level it was necessary to add an unlabelled "carrier" protein (e.g. 5mg sperm-whale apomyoglobin) as the low amounts of protein are otherwise extracted from the cup in the organic solvents. A similar finding has recently been reported by others (McKean et al, 1974; Niall et al, 1974). For rat Ig L-chains 20 mg was the upper limit - with larger amounts the protein flaked off the cup during drying steps.

For application the sample was dissolved in about 0.5 mls of a volatile solvent and dried by application of vacuum to the spinning cup. Water or 0.1M ammonia solution were generally preferred, but formic acid was used occasionally when the former failed to dissolve the sample. Solution of the sample, however, for application or during the degradation was not essential for efficient reaction. In cases where the sample was applied as a suspension, or where the R2 or R3 failed to dissolve the film, the success of the run appeared unaffected.

In some early runs a problem of detachment of the protein film during drying stages was encountered. This was later avoided by routine pretreatment of the protein film by solution in heptafluorobutyric acid and washing with chlorobutane according to the last part of the

programme.

3.6 Programmes

The programme most commonly used for degradation of proteins ("Fast Protein, Quadrol 072172") is available from Beckman, and is closely based on the original programme (Edman & Begg, 1967). In outline the stages are:

1) Coupling (12 mins). Reaction with phenylisothiocyanate.
2) Benzene Wash (3 mins). Removal of diphenylthiourea and other byproducts.
3) Ethyl acetate wash (10 min). Removal of Quadrol.
4) Fine Vacuum. Drying of protein film.
5) First cleavage (3 mins). Heptafluorobutyric acid.
6) First extraction (3 mins). 1-chlorobutane containing cleaved aminoacid derivative to fraction collector.
7) Second cleavage. Repeat of (5) in case reaction did not go to completion.
8) Second extraction. Repeat of (6), but deliver to waste.
9) Fine Vacuum. Drying of protein film to remove all acid.

"Fine Vacuum" is always preceded by "Restricted vacuum" in which gentle vacuum is applied via a "Teflon" capillary followed by "Rough Vacuum" which uses a more highly ballasted pump than the final "Fine Vacuum". Only those prolonged vacuum stages which result in a dry protein film are indicated. Short vacuum stages are also used to withdraw reagents and solvents from the delivery lines and as a preliminary removal of volatile material from the cup.

3.7 Conversion of fractions

The first chlorobutane wash from each cycle is collected in a refrigerated fraction collector, dried in a stream of nitrogen and kept under vacuum. The amino acid derivative at this stage (anilino-

thiazolinone) is converted to the more stable phenyl thiohydantoin (PTH) by adding 0.2 mls N HCl and heating to 80°C for 10 mins in an oil bath. To reduce destruction by oxidation, especially of the serine and threonine derivatives, each tube is flushed with nitrogen after the addition of HCl, and stoppered with a silicon rubber bung.

3.8 Extraction of fractions

After conversion the samples are cooled and extracted twice with 0.7 mls ethyl acetate. The organic phases, which contain all PTH-amino acids except PTH-His, PTH-Arg and PTH-Cys (O_3H), are evaporated to dryness.

The conversion and extraction are generally carried out on 12 samples simultaneously, enabling the total daily output of the machine to be processed in less than one hour. There is a danger of introducing sequence errors at this stage through mishandling of tubes. To avoid this, scrupulous care was taken to maintain a strict routine. Each tube was labelled with a serial number identifying the run and cycle before being placed in the fraction collector and this was checked on removal of the tube. From then on each operation was carried out in order of serial number to reduce further the chance of an error.

Wittmann-Liebold (1973b) has developed an automatic conversion device which eliminates the possibility of error at this stage. Conversion is achieved with aqueous trifluoracetic acid (20% v/v) at 55°C for 30 min. Most of the reagents and solvents used by her are repurified and of sufficient purity to permit direct identification of the PTH-derivatives (by TLC) without ethyl acetate extraction. It is possible that the immediate conversion of the thiazolinones has the further advantage of more reproducible yields of certain amino acid derivatives.

3.9 Identification of PTH-derivatives

Thus far there is no single technique for the satisfactory identification of all PTH-amino acids. A variety of methods is available and most laboratories use a selection which between them cover all likely PTH-amino acids.

3.9.1 Thin layer chromatography (TLC)

In their original work Edman and Begg (1967) used TLC on silica to determine the first 60 residues of hump-back whale myoglobin, except for PTH-Arg and PTH-His which were identified by thin-layer electrophoresis on the same support. The method is rapid and simple to use, requiring no expensive equipment. Because of these advantages Edman (1970, and personal communication) has continued to use TLC as the principal method for PTH-amino acid identification in his further development of the sequenator. An apparatus has been described for the conversion, extraction and application of PTH-amino acids in batches of ten, thus reducing the risk of accidental exchange of tubes.

Other laboratories have also relied mainly on TLC (e.g. Wittmann-Liebold, 1973a). However the method suffers from a major disadvantage in that it is not possible to quantitate the yield of each derivative in a mixture. As the run progresses, a "background" of PTH-amino acids develops due to non-specific cleavage of the polypeptide chain on exposure to acid. When the level of this "noise" approaches that of the steadily decreasing "signal", it becomes difficult to distinguish signal from noise. In these studies TLC has been used to complement more quantitative methods and only for that part of each run where the background is not yet strongly visible (generally the first 15-25 residues).

The details of the method are as follows:

Aliquots (5-10 nmoles) from the ethyl acetate extracts are spotted

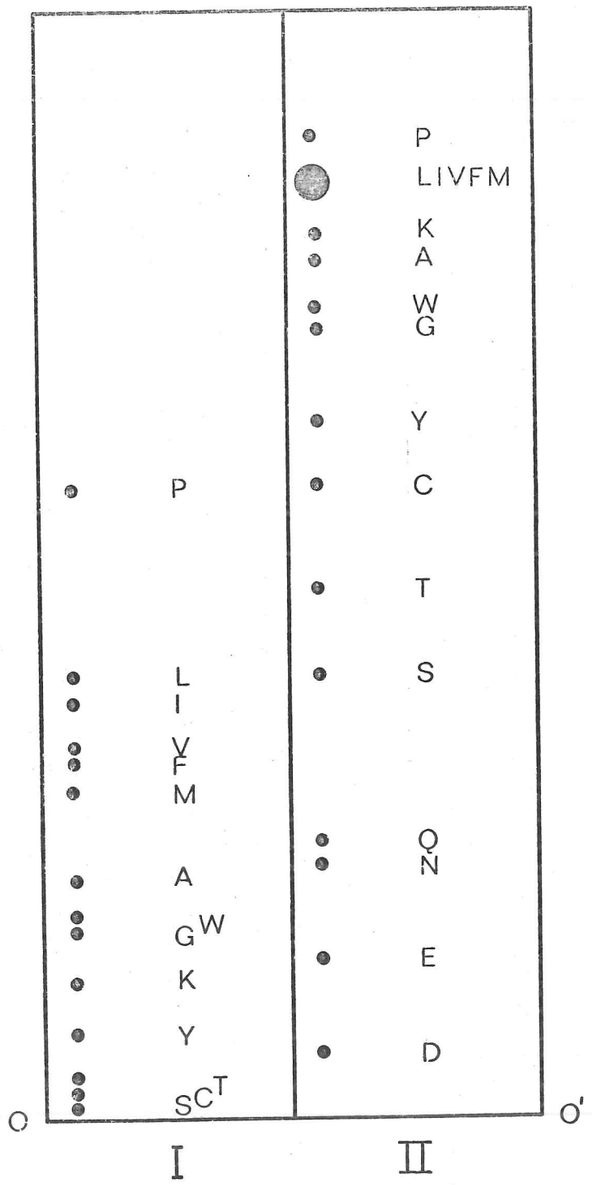


Fig. 3.4 Ascending thin layer chromatography of PTH-amino acids on silica. The movement of the solvent front is towards the top of the page. OO' is the origin. I) Solvent system I ($CHCl_3/C_2H_5OH$, 98.2); II) Solvent system II ($CHCl_3/CH_3OH$, 9:1). In solvent I PTH-Arg, -His, -Asn, -Asp, -Gln, -Glu remain at the origin and are not shown. In solvent II PTH-Arg and PTH-His remain at the origin and these are the only derivatives not resolved by the combination of both solvents. However these two PTH-amino acids are not encountered in the normal identification procedure as they remain in the aqueous phase.

onto duplicate 20cm x 20cm silica thin layer plates which contain a fluorescent indicator (60F 254, Merck, Darmstadt, Germany). PTH-reference mixtures are also applied. Both plates are developed for about 1 hr by ascending chromatography in pre-equilibrated glass tanks at room temperature. The solvent for one plate is $CHCl_3/C_2H_5OH$ (98:2 (v/v)) (Solvent I, Terhorst et al, 1973); the other solvent is $CHCl_3/CH_3OH$ (9:1) (Solvent II, Terhorst et al, 1973). After development the plates are dried and inspected under UV light (254 nm). Since the intensity of the spots (especially PTH-Ser and PTH-Thr) fades rapidly a photographic record is made on Tri-X film (Kodak Ltd., Stevenage, UK) exposed through a X8 red filter.

Solvent I resolves the PTH-derivatives of Pro, Leu, Ile, Val, Phe, Met, Ala, Trp, Gly, Lys, Tyr; Solvent II resolves those of Tyr, Thr, Ser, Gln, Asn, Glu and Asp (Figs. 3.4). PTH-His and Arg, which in any case remain in the aqueous phase after extraction, remain at the origin after development in both solvents.

3.9.2 Gas Liquid Chromatography (GLC)

GLC is the most widely used identification method for PTH-amino acid identification. Most amino acid derivatives can be analysed at the nanomole level and in the case of the hydrophobic amino acid derivatives the data is quantitative. The method used in the studies described in this and the following chapter is a modified version of the procedure of Pisano and Brozert (1969).

A Beckman GC-65 gas chromatograph containing two "U" glass columns (4 ft x 2 mm i.d.) in a single oven was used. The columns were used for independant simultaneous analysis of two samples. The second column can alternatively be used as a reference column but with careful conditioning (see below) this was unnecessary. The column filling was Chromosorb W (Phase-Separations Ltd., Queensferry, UK)

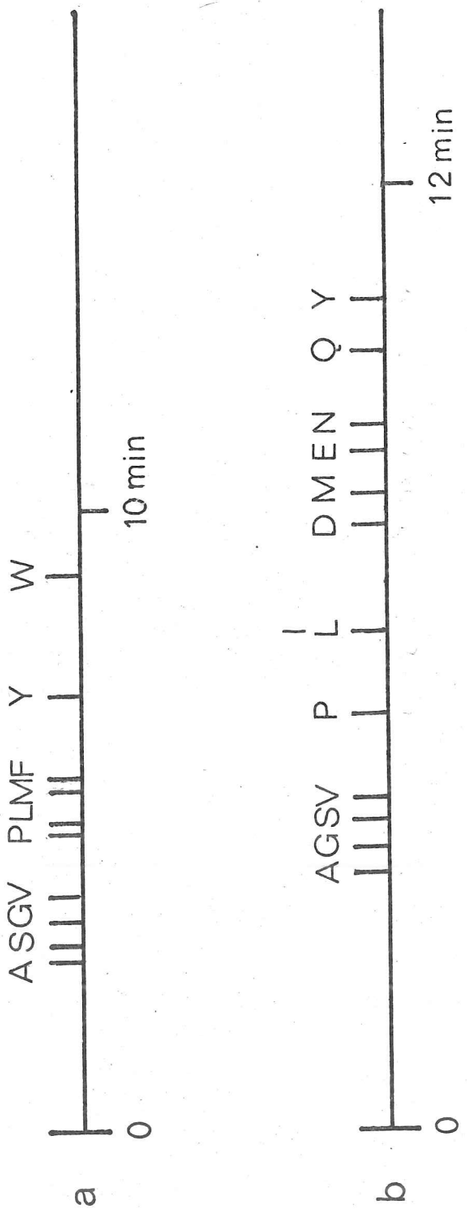


Fig. 3.5 Order of elution of PTH-amino acids from SP400-coated GLC columns (120cm x 2mm). Where two derivatives elute together they are both shown. a) "Direct", 1 µl of PTH-amino acid solution injected at time 0; b) "silylated", 1 µl of PTH-amino acid solution injected together with 1 µl BSA. PTH-Ile elutes as a doublet peak after silylation and may thus be distinguished from PTH-Leu.

coated with 10% SP400 (Supelco Inc., Bellefonte, Pa., USA) (Hermodson et al, 1972) or as supplied (Beckman) and was "conditioned" by heating from 50°C to 325°C over a period of 5 hours, followed by 16 hours at 325°C. The same temperature programme was used for the elution of both PTH-amino acids and of the silylated derivatives (see below). It consisted of a 5 min isothermal at 190°C, a linear rise of 20°C/min to 280°C and a second isothermal for 5 min. The complete cycle time, including automatic return of the instrument to the starting conditions was 15 min., a considerable saving on the standard 30 min programme (Hermodson et al, 1972). The resolution of PTH-amino acids was maintained by increasing the carrier (He) gas flow rate from 60 to 100 mls/min. Samples (usually 1-5 nmoles in 1 or 2 µl ethyl acetate) were injected from a Hamilton Syringe. Detection of PTH-amino acids was by flame ionisation in a hydrogen/air/helium flame (40:50:300 ml /min). The PTH-derivatives of Ala, Gly, Val, Met, Phe, Tyr, Trp were detected unambiguously in this way. PTH-Ser/PTH-CMCys eluted together as did PTH-Pro/PTH-Thr and PTH-Leu/PTH-Ile. The remaining derivatives are either destroyed or not sufficiently volatile. A second aliquot of each sample was submitted to GLC (under the same conditions) after derivatisation with N,N-bis-trimethyl-silyl acetamide (BSA, BDH Ltd.). The reaction was carried out "on-column" by injecting 1 µl of BSA reagent "sandwiched" in the syringe between 1 µl of ethyl acetate and the sample, which is injected first. The "silylated" derivatives of Thr, Pro, Leu, and Ile (which forms a doublet peak) are now unambiguously separated and also the derivatives of Asp, Asn, Glu and Gln which are sufficiently volatile. For each GLC peak the time of elution and the magnitude of the response differed slightly between the two columns. For this reason one column was reserved for the silylated derivatives of a given series and the other for the unsilylated PTH-amino acids. In this way comparisons of yields of various derivatives in adjacent

fractions was made easier.

3.9.3 Liquid scintillation counting

To distinguish PTH-CMCys from PTH-Ser, the carboxymethylation was performed with iodo-2-^{14}C-acetic acid and an aliquot of each converted fraction was removed and counted in Bray's solution (see Section 2.18). Liquid scintillation counting was also used to detect other "external" radio-labels (e.g. ^{3}H-dinitophenyl-lysine) and "internally" labelled residues (e.g. ^{35}S-methionine) as described below (Section 3.13.2).

3.9.4 Hydrolysis of PTH-amino acids

"Back hydrolysis" of PTH-amino acids to regenerate the free amino acid was investigated by Van Orden and Carpenter (1964). They found that 0.1N NaOH gave high yields of most amino acids, which can then be analysed in the normal way using conventional amino acid analysers. 6N HCl at 150^{o}C for 48 hr gave inferior recoveries owing to the slow rate of hydrolysis. PTH-Ser is however completely destroyed under the alkaline conditions and PTH-Thr converted to Gly with 64% yield (Africa and Carpenter, 1966). Hydriodic acid (HI) had been suggested as a more rapid hydrolysis medium for proteins by Baernstein (1936) and investigated further by Inglis *et al* (1971). During HI hydrolysis PTH-Met and PTH-Trp are destroyed, PTH-Ser converted to PTH-Ala and PTH-Thr to α-amino butyric acid. PTH-Gln and PTH-Asn are hydrolysed to the corresponding free acid forms. Smithies *et al* (1971) have used the combination of HI and NaOH hydrolysis as the sole means of PTH identification. This has the advantage that since the thiazolinone derivatives hydrolyse to amino acids equally well, the conversion step may be omitted and since the amino acid analyser is less sensitive to non-amino acid impurities, extraction

Amino acid	% yield
Glu	86.9
Leu	82.6
Asn	81.4
Pro	80.7
Phe	80.2
Ile	76.5
Ala	73.7
Tyr	72.3
Asp	70.5
Lys	70.3
Val	69.8
Thr	57.1
[a]Arg	50.0
His	49.2
Gln	47.3
Ser	32.4

Table 3.1. Recoveries of free amino acids after hydrolysis of PTH-amino acids with HI. 25 nmole samples of each PTH-amino acid were hydrolysed. Yields are expressed relative to the yield of norleucine from PTH-norleucine which was added to each sample as an internal standard. The mean value of triplicate hydrolyses is shown. Reproducibility was within 10%. The mean value of norleucine recovery was 42.6% (±3.9%) and norleucine recoveries were all normalised to this value. PTH-Met and PTH-Trp give irreproducible and/or unidentified breakdown products. PTH-CMCys gives a variable yield of Ala. PTH-Thr yields α-amino butyric acid, PTH-Ser yields Ala and PTH-Ile a mixture of Ile and allo-Ile. a) Tentative value derived from a single hydrolysis. Amides were identified as the free acids.

is also unnecessary.

In the studies described here HI hydrolysis was used to supplement other means of identification. In general this was carried out as follows:

2%-4% of the ethyl acetate extract was subjected to GLC and a further 2%-4% after silylation. 10-20% was taken for TLC. At this stage any sample which had not been positively identified by both GLC and TLC was reconstituted with its aqueous phase and dried in a 10mm x 75mm pyrex hydrolysis tube. 200 µl of 55% HI (B.D.H. Chemicals Ltd.) was added from a freshly opened vial. The tube was flushed with nitrogen and sealed under vacuum. After hydrolysis for 24 hr at 130°C the tubes were opened and dried over NaOH. A suitable amount (usually 25%-100%) of each hydrolysate was applied to an amino acid analyser. PTH-amino acid standards were processed in a similar fashion to determine a destruction constant for each PTH-derivative (Table 3.1).

HI hydrolysis complements well GLC for PTH-amino acid identification. Those PTH-amino acids which are destroyed by HI (Met, Trp) are easily (and quantitatively) detected by GLC. The identification of PTH-Thr as α-amino butyric acid is entirely satisfactory as this amino acid elutes between Ala and Val on the first (pH 3.25) buffer. The identification of PTH-Ser after reduction to Ala is in general also satisfactory. PTH-Ser and PTH-Ala are resolved by GLC and TLC and the partial destruction of PTH-Ser results in a lower yield of Ala than would be expected from PTH-Ala. Unfortunately PTH-CMCys is also hydrolysed to Ala by HI and thus sequences which are rich in Ser, CMCys and Ala (e.g. rat kappa chains residues 20-30) present special difficulties.

The disadvantages of hydrolysis are:

1) the loss of information about amides

2) the labour and capital equipment required-a conventional amino

acid analyser is required full-time for the analysis of a sequenator's output.

3) the delay between obtaining samples from the sequenator and completing the amino acid analysis (often one week) means that the method is not suitable for monitoring the progress of a run with a view to deciding, for example, whether to continue or to stop the run, or whether to modify the programme.

3.9.5 Spot tests

PTH-Arg may be identified by the phenanthraquinone reaction (Easley et al, 1969) or a modified Sakaguchi reaction (Acher & Crocker, 1952). PTH-His may similarly be identified by the Pauli test (Easley, 1965) or by the method of Sanger and Tuppy (1951). However these reactions are not very sensitive and false positives have been reported (Niall, 1973). In the later stages of a long degradation the background may obscure the identification of a weakly positive residue. For these reasons and since, in the proteins studied here, His and Arg did not occur very commonly and were therefore easily identified by back hydrolysis, such spot tests were not routinely used.

3.9.6 Mass spectrometry

It has been reported that all PTH-amino acids may be identified by mass spectrometry (Weygand & Obermeier, 1971). The method is rapid and quantitative but the high capital expenditure and expert maintenance required for the equipment has deterred most laboratories.

3.9.7 High pressure liquid chromatography (HPLC)

This technique is still at the development stage, but preliminary results have been reported (Zimmerman et al, 1973; Frank & Strubert, 1973; Graffeo et al, 1973). Detection of PTH

derivatives by their absorbance at 254nm is sensitive and could be improved even further by the use of fluorescence detectors. The preparation of samples is minimal and the loading easily automated. However a satisfactory solid support for single column operation has yet to be discovered. Reports that a two-column system resolves all PTH-amino acids (Appella, personal communication) awaits confirmation, using sequenator samples of its practical appliability.

3.10 Interpretation of Data

The rapid expansion in the number of sequenators and the corresponding increase in the sequence data published has led to a certain amount of questioning of the accuracy of the sequences thus determined. Presentation of the yield of PTH-amino acid released at each cycle of the degradation expressed as the exponential decline in yield with cycle number has been presented (e.g. Bennich et al, 1973; Kolb et al, 1974) but this does not in general reflect the way in which the sequence was in fact determined. With the combination of use of GLC, TLC and amino acid analysis, the identification of the sequence residue is usually semi-quantitative and depends on which (and how many) of the three methods was used as well as on the particular PTH-amino acid.

The repetitive yield (i.e. the average yield per cycle of degradation) is not a good guide to the confidence of a particular sequence. The repetitive yield which is a guide to the performance of the machine and to the progress of a run, is calculated from data on those PTH-amino acids which are the easiest to identify and quantitate. The problems of identification of sequence residues however occur first with those residues whose identification is more difficult and only semi-quantitative. Edman (1970) has alluded to the difficulty in presentation of such data, but no satisfactory solution has been suggested and indeed even in the original paper of Edman and Begg (1967) only part of the

data is presented. Now that more techniques are generally used for PTH-amino acid identification the volume of primary data is even more prohibitive to its entire presentation and most sequences determined automatically are now published without any primary data (e.g. Ramshaw et al, 1974; Koo & Cebra, 1974).

To illustrate the process of sequence identification used in these studies the data of part of a run described in the following chapter is shown in Figs. 3.4 - 3.8 and discussed below.

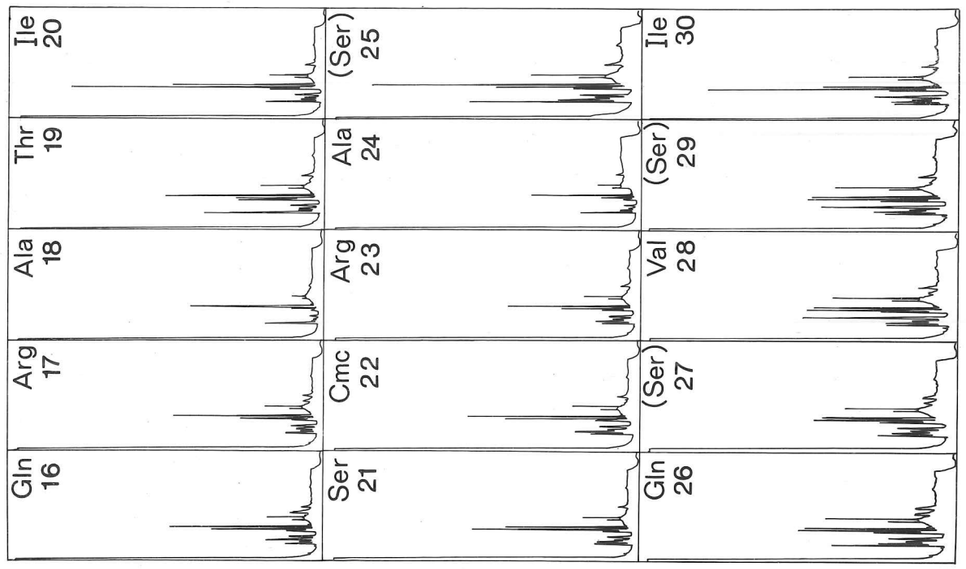
3.11 Internal standard

To improve the accuracy of quantitative calculations PTH-norleucine was added to each of the fraction collector tubes before beginning a run. The PTH-norleucine is then carried through the same processes of conversion, extraction and identification along with the unknown PTH-amino acids. PTH-norleucine does not interfere with the identification of any other PTH-amino acid in TLC or GLC and norleucine is eluted between Leu and Tyr on the second buffer(pH 4.25) of the amino acid analyser. Quantitation of the internal standard served as a check on any losses between delivery to the fraction collector and identification and also for any variation in sampling for identification (except for those PTH-amino acids that remain in the aqueous phase after extraction and thus are not removed in the same proportion as the internal standard).

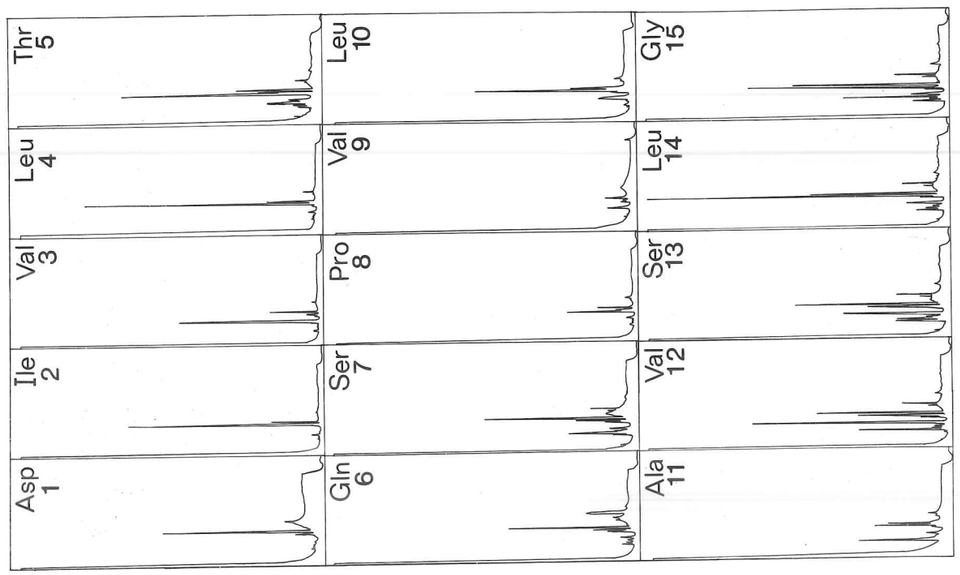
3.12 Some notes on sequence identification

IR97 (see chapter 4) was selected, as a typical run, to illustrate the way in which the assignment of sequence residues was made. The first 30 residues are discussed.

Fractions 1-25 contained 50 nmoles, 26-30 contain 25 nmoles PTH-norleucine as internal standard. After conversion and extraction,



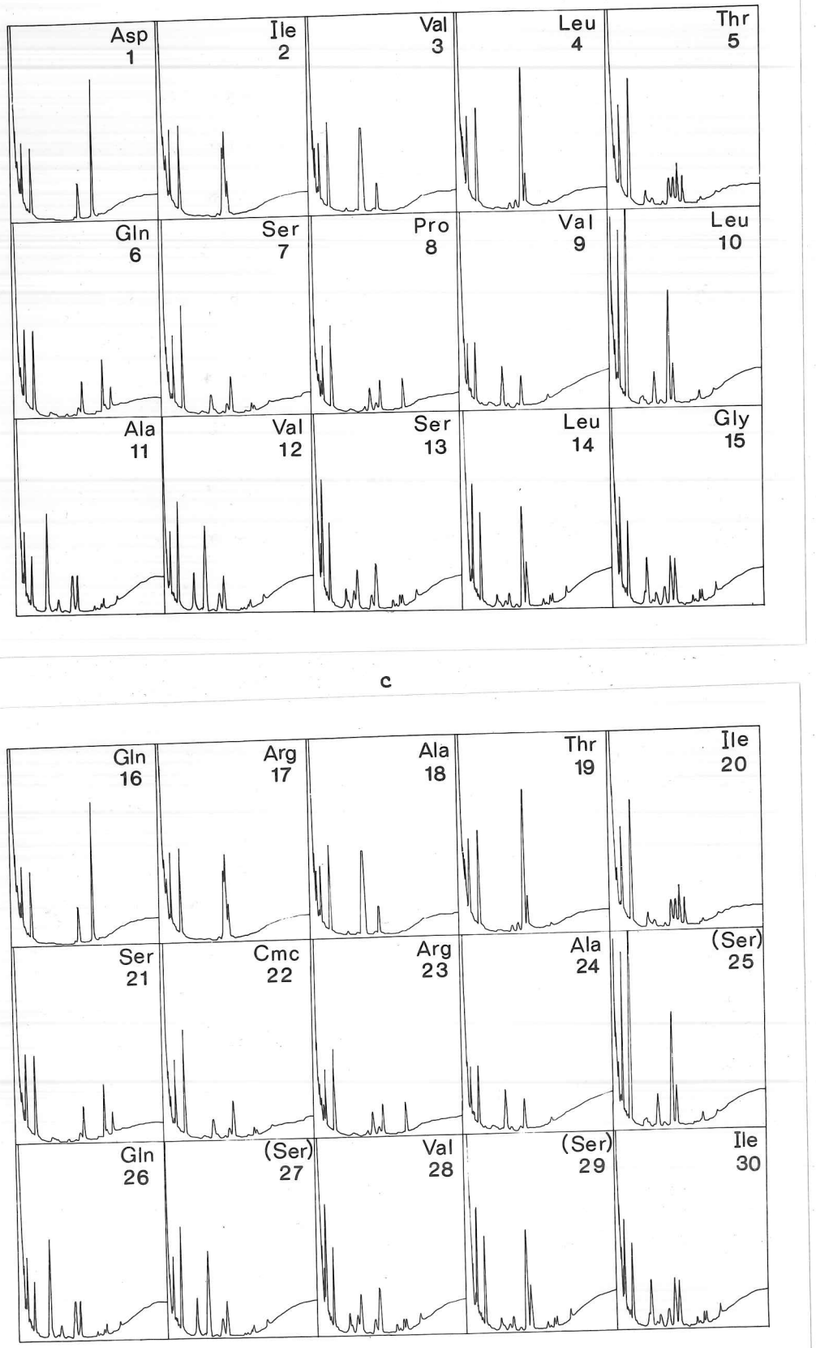
b



a

Fig. 3.6 GLC analysis of sequenator run on IR97. a) Fractions 1-15, Direct; b) fractions 16-30, Direct; c) fractions 1-15, silylated; d) fractions 16-30 silylated. The fraction number and final amino acid assignment (not necessarily derived from GLC alone) are shown together with each chromatogram. Brackets indicate a tentative assignment (see text). The numbering does not correspond to that used in chapter 4 where a deletion has been assigned to position 10. Fractions 1-25 contain 50 nmoles of internal standard (PTH-norleucine), fractions 26-30 contain 25 nmoles. Other variation in the peak height of the internal standard reflects changes in electronic amplification of the detector signal.

Fig. 3.6. For legend see facing page.



c

d

Fig. 3.6 (cont'd.).

the PTH-amino acids were dissolved in 50 µl ethyl acetate. 1 µl of this solution was injected directly into the gas chromatograph ("Direct GLC"), 1 µl injected with 1 µl BSA, 5 µl (residues 1-22) or 15 µl (residues 23-30) applied to TLC plates, 5 µl counted for radioactivity and the remainder pooled with the aqueous phase and hydrolysed with HI.

Residue 1 Direct GLC: no major peak (other than that due to PTH-norleucine); after silylation: a single major peak with elution time corresponding to "silylated" PTH-Asp; after TLC in solvent I: a single spot at the origin was visible. In solvent II this moved with the mobility of PTH-Asp. No trace of PTH-Asn was visible in GLC or TLC.

Residue 2 Direct GLC: a single major peak in the position of PTH-Leu/-Ile; after silylation a doublet peak characteristic of PTH-Ile rather than PTH-Leu. TLC also identified the derivative as that of Ile.

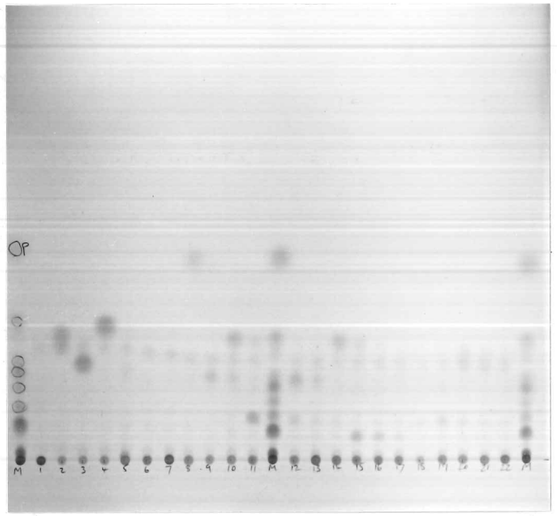
Residue 3 PTH-Val was identified by its elution times on GLC before and after silylation and also by TLC.

Residue 4 Direct GLC: PTH-Leu/-Ile (as residue 2); silylation produced a single PTH-Leu peak; TLC confirmed PTH-Leu.

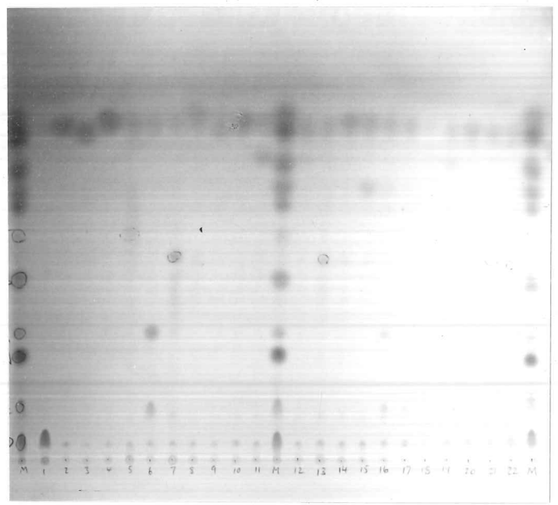
Residue 5 Direct GLC: major peak PTH-Pro/-Thr, minor peak PTH-Gly which is a characteristic breakdown product of PTH-Thr; TLC: weak PTH-Thr spot after solvent II.

Residue 6 Direct GLC: no major peak, minor peak of PTH-Gln; after silylation: major peak PTH-Glu but also (~ 30%) PTH-Gln; TLC (solvent II): major spot PTH-Gln, also PTH-Glu.

Residue 7 Direct GLC: major peak PTH-Ser (10 x background); after silylation: PTH-Ser; TLC: PTH-Ser major spot after solvent II.



a



b

Fig. 3.7 TLC analysis of sequenator run on IR97. 5µl (10%) of each ethyl acetate extract was applied to each plate. Only fractions 1-22 are shown; the background in later fractions was too high to permit reliable identification. a) After development in solvent I; b) after development in solvent II. PTH-Ser and PTH-Thr spots were outlined in pencil because they fade rapidly. Plates were photographed under uv illumination (254 nm). For final assignments of each fraction see Fig. 3.6. The reference mixture of PTH-amino acids (M) contains the derivatives of Pro, Leu, Val, Phe, Ala, Gly, Thr, Ser, Gln, Asn, Glu, Asp.

Residue 8 Direct GLC: PTH-Pro/-Thr; after silylation: PTH-Pro; TLC: PTH-Pro after solvent I.

Residue 9 Direct GLC: PTH-Val; after silylation: PTH-Val; TLC: PTH-Val after solvent I.

Residue 10 PTH-Leu identified as at residue 4. (Repetitive yield for 4-10 = 91.6%).

Residue 11 Direct GLC: PTH-Ala; after silylation: PTH-Ala; TLC (solvent I): PTH-Ala (weak spot PTH-Leu due to overlap).

Residue 12 Direct GLC: PTH-Val; after silylation: PTH-Val; TLC (solvent I): PTH-Val (Repetitive yield for 3-12 = 94.1%).

Residue 13 Direct GLC: PTH-Ser (5 x background, some PTH-Val overlap); after silylation: PTH-Ser; TLC (solvent II): PTH-Ser is major spot.

Residue 14 PTH-Leu identified as residue 4. 4 x background on direct GLC. Repetitive yield 4-14 = 92.7%.

Residue 15 Direct GLC: PTH-Gly (15 x background); after silylation: PTH-Gly; TLC = PTH-Gly major spot, also some PTH-Leu overlap.

Residue 16 Direct GLC: PTH-Gln minor peak; after silylation: PTH-Glu (5 x background) and PTH-Gln (∿ 30% yield of PTH-Glu); TLC (solvent II): major spot PTH-Gln, also PTH-Gln. After HI hydrolysis major peak is Glu.

Residue 17 No GLC peak significantly above background; weak unidentified TLC spot remains near origin after solvents I and II. Suggests His or Arg. After HI hydrolysis major amino acid is Arg.

Residue 18 Direct GLC: PTH-Ala (5 x background); after silylation: PTH-Ala; TLC: no major spot after solvents I and II; PTH-Ala minor spot after solvent I. (Low yield of background and sample but not of internal standard suggests loss of sample between cleavage in cup and delivery to fraction collector).

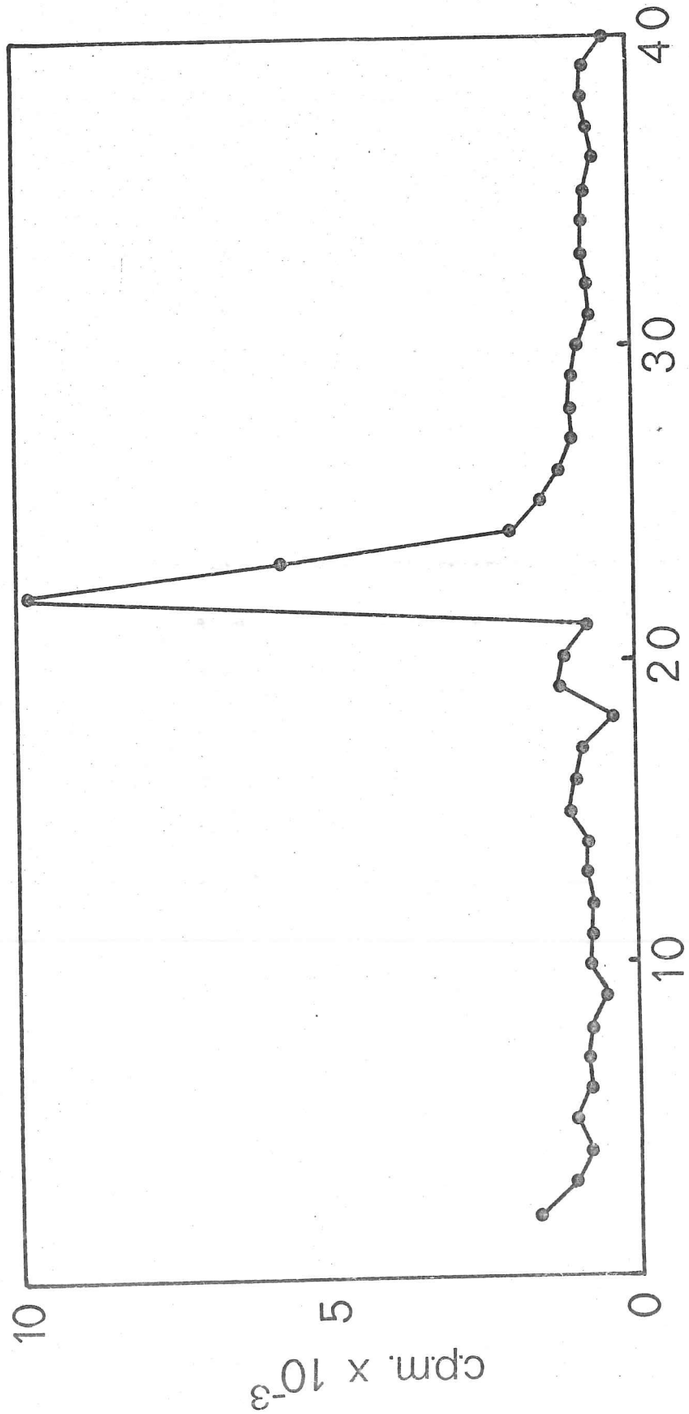


Fig. 3.8 Identification of radiolabelled carboxymethyl cysteine by liquid scintillation counting of fractions 1-40 of sequenator run on IR97 (see section 3.9.3). 5 µl (10%) of the ethyl acetate extract was counted in Bray's solution. The c.p.m. shown are the uncorrected values.

Residue 19 Direct GLC: PTH-Pro/-Thr (3 x background), some PTH-Ala overlap confirms Ala at 18; after silylation: some PTH-Gly (breakdown product of PTH-Thr); TLC (solvent I) PTH-Ala (overlap); solvent II: weak PTH-Thr; HI hydrolysis: major peak α-amino butyric acid confirms Thr.

Residue 20 PTH-Ile identified as residue 2, but also confirmed after HI hydrolysis. (Repetitive yield for 2-20 = 95.2%).

Residue 21 Direct GLC: PTH-Ser is 2 x background; PTH-Ser also identified after silylation and after TLC in solvent II.

Residue 22 Liquid scintillation counting shows ^{14}CM-Cys at this position and at no other position Direct GLC: peak at position of PTH-Ser is PTH-CMCys.

Residue 23 PTH-Arg identified as residue 17.

Residue 24 PTH-Ala identified as residue 11. Direct GLC: PTH-Ala is 4 x background. (Repetitive yield for 11-24 is 92%).

Residue 25 Direct GLC: PTH-Ser peak elevated, but less than 2 x background. Some overspill from PTH-Ala (overlap from residue 24) cannot be excluded. Yields were too low and background too high for TLC to be of regular further use. After HI hydrolysis there was insufficient Ala produced, relative to the high Ala "noise", to confirm the assignment of Ser. The quantitation of Ala was complicated in this case by the presence of Ala overlap from residue 24 as well as Ala background from PTH-Ser and PTH-Ala. All PTH-amino acids other than PTH-Ser would, however, have been detected by GLC or amino acid analysis and thus PTH-Ser is also suggested by elimination of other possibilities. However in the absence of confirmatory evidence the assignment of Ser was described as tentative.

Residue 26 Direct GLC: peak of PTH-Gln; after silylation: PTH-Glu and PTH-Gln; after HI hydrolysis: Glu.

Residue 27 In the absence of confirmatory evidence Ser was tentatively assigned to this position (see residue 25).

Residue 28 Direct GLC: PTH-Val (2.5 x background); after silylation: PTH-Val; after HI hydrolysis Val.

Residue 29 Direct GLC: PTH-Ser (1.3 x background), after silylation: PTH-Ser weakly positive; after HI hydrolysis, Ala is 1.2 x background.

Residue 30 Direct GLC: PTH-Leu/-Ile (1.8 x background) after silylation PTH-Ile; after HI hydrolysis: Ile (Repetitive yield 2-30 = 94.5%).

3.13 Applications

As a demonstration of sequenator's potential Edman & Begg (1967) determined the first 60 amino-terminal residues of hump-back whale apomyoglobin. The early applications to which commercial sequenators were put consisted mainly of the determination of unknown N-terminal sequences, both as a preliminary to total sequence determination (e.g. Li & Riggs, 1970; Imamura et al, 1972) and as a characterisation in itself of individual proteins (e.g. Benditt et al, 1971; Wang et al, 1970).

The scope of the instrument has been extended in a number of ways by its application to sequence determination of protein fragments.

Two general approaches to the sequencing of fragments may be distinguished. That of Braunitzer has been to follow the classical strategy of cleaving the protein into fragments, principally using trypsin, and to modify chemically the resulting small peptides so that they are less soluble in the organic extraction solvents. The development of a series of reagents to modify the ε-amino group of lysine (Braunitzer et al, 1971) has culminated in the successful total sequence determination of bovine β-lactoglobulin exclusively by automatic sequencing (Braunitzer

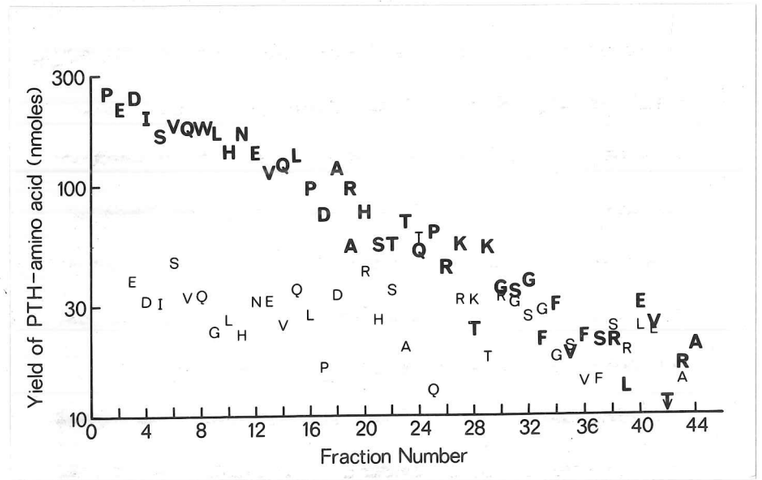


Fig. 3.9 Sequenator yields from ε2 fragment. 500 nmoles of the CNBr fragment was degraded with the standard (protein) programme. Bold characters indicate the sequence residue assigned to each position. Light characters the major impurity.

et al, 1973).

The second approach has been to try to obtain large fragments (usually from CNBr cleavage) and to attempt long degradations on these directly without further cleavage (Niall et al, 1970). The volatile buffer dimethylallylamine (DMAA) has been much used in this approach, as an alternative to Quadrol, in order to combat extractive losses as the fragment becomes smaller and more soluble in the organic solvents (Niall et al, 1969). Dimethylbenzylamine, which is involatile in the sequenator, but soluble in benzene, has also been used to avoid the necessity for long ethylacetate extractions (Hermodson et al, 1972).

A detailed scheme for the sequencing of peptides by successive modifications of the programme and buffers has been suggested by Niall (1971) and recently applied (Kolb et al, 1974).

The following section describes the automatic sequencing of two immunoglobulin fragments which served as a guide for the use of some of the modifications described above.

3.13.1 IgE fragments

Two CNBr fragments prepared from the H-chain of the human myeloma protein ND (Bennich & Johansson, 1971) were subjected to automatic degradation. The fragments were provided by Dr H. Bennich.

The standard double cleavage programme was used for 50 cycles on the longer fragment, ε2. The results of this run are summarised in Fig. 3.9. The repetitive yield was about 94% and the fact that the graph is approximately linear throughout the run suggests that even when the fragment remaining in the cup contained only 35-40 residues it was not subject to appreciable "wash-out". The two arginine residues towards the C-terminus (Bennich et al, 1973) are no doubt largely responsible for this. Owing to the size of the fragment (77 residues) the background remained low throughout the run. Thus the "overlap" sequence could be followed easily and this served as an independant check on the sequence.

Table 3.2

Sequencer yields of ε4-fragment

Fraction	Major		Minor**	
	Residue	(nanomoles)	Residue	(nanomoles)
1	Arg	400	-	-
2	Ser	>400*	-	-
3	Thr	>170*	-	-
4	Thr	>170*	-	-
5	Lys	>200*	-	-
6	Thr	100	Lys	32
7	Ser	370	-	-
8	Gly	>310*	-	-
9	Pro	210	-	-
10	Arg	92	Pro	90
11	Ala	200	Arg	44
12	Ala	> 50		

* Peak off-scale on amino-acid analyser

** No values given if the yield was lower than 20% of main residue

This application also illustrated the way in which automatic and manual data can complement each other. Much of the sequence of ε2 had previously been obtained using the dansyl-Edman procedure. However certain overlaps near the amino-terminus were missing and also a phenylalanine-rich region (30-37 residues from the amino-terminus of ε2) which could not be obtained in good yield from enzymic digests. Both these problems were solved easily with a single sequenator run. The results from the sequenator, which were obtained quite independantly of earlier data also served to confirm the manual sequence determination.

A second, smaller (40 residue) fragment, ε4, from the same H-chain was also applied to the sequenator. In an attempt to reduce extractive losses, a single cleavage programme was used; the Quadrol buffer diluted to 0.25 M with propan-2- ol/water (3:2, v/v); the coupling time increased to 30 min (from 12 min) and the ethyl acetate wash reduced to 400 sec (from 600 sec). The sequence Arg-Ser-Thr-Thr-Lys-Thr-Ser-Gly-Pro-Arg-Ala-Ala was obtained (Table 3.2) and provided the desired overlaps. Had it been necessary to extend this sequence the use of DMAA and/or modification of the remaining lysine with a Braunitzer reagent would have been attempted on a second sample.

3.13.2 Use of Radioactivity

The use of radioactivity allows a dramatic increase in the sensitivity of the automatic degradation procedure and is likely to become the next major development in the sequenator methodology. Proteins may be labelled "externally" - usually by specific chemical modification (e.g. iodo-2-^{14}C-acetic acid)- butalso generally by ^{3}H exchange (Hembree et al, 1973; Ghanem & Westermark, 1960; Gosztonyi & Walde, 1966) or "internally" by synthesis of the protein, in vivo or in vitro, in the presence of labelled amino acids. Radioactivity may also be incorporated in the sequenator by using radioactive PITC as suggested by Laursen (1969)

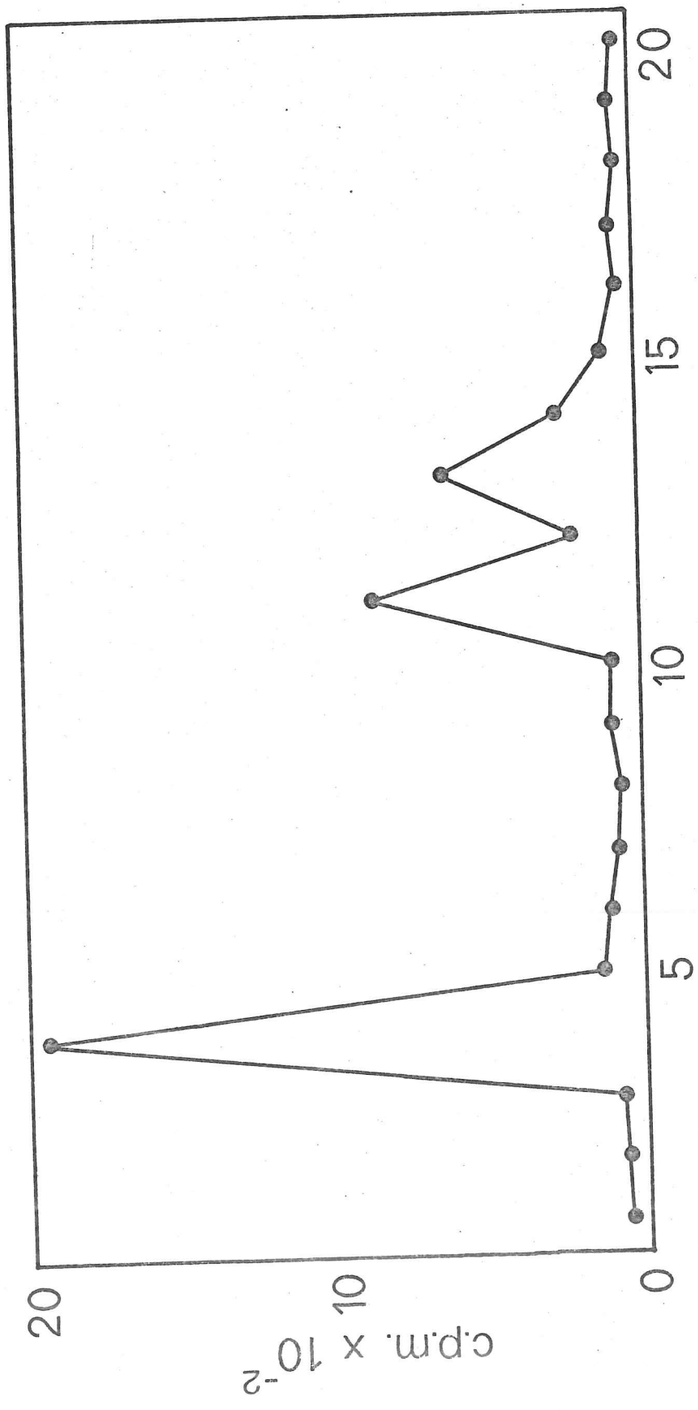


Fig. 3.10 Characterisation of P3K L-chain by identification of Met at positions 4, 11, 13. 46,000 c.p.m. of ^{35}S-Met-labelled P3K L-chain, +5mg sperm whale apomyoglobin as carrier, was subjected to 20 cycles of sequenator degradation. 25% of each fraction was removed for counting.

for solid phase sequencing. Whichever method is used, if only one amino acid is known to be labelled, no separation of the PTH derivatives is required before counting. However the sequence information which can be derived from such a run by radioactivity measurements is clearly limited. Where two or more labelled precursors are used, partial or total separation of the PTH-amino acids is required before counting. Of course, the use of different isotopes which can be resolved (e.g. ^{3}H and ^{35}S) increases the scope of this approach.

In these studies, apart from the routine use of iodo-2-^{14}C-acetic acid described above, the application of radioactive methods was explored in two types of experiment.

3.13.2.1 P3K L-chain - P3K cells were grown in the presence of ^{35}S-methionine and the secreted Ig prepared as described in Chapter 2. After partial reduction and carboxymethylation, H- and L-chains were separated on Sephadex G-100.46,000 cpm of ^{35}S-Met P3K L-chain was applied to the sequenator, together with 5 mg sperm whale apomyoglobin as carrier, and degraded for 19 cycles. After conversion of each fraction 50 µl (from 200 µl) was removed for liquid scintillation counting. The results are shown in Fig.3.10which shows that radioactivity was released from the protein mainly in cycles 4, 11 and 13. From a knowledge of the L-chain sequence (Svasti & Milstein, 1972b) these are the positions of methionine in the first 19 residues, confirming that the degradation proceeded correctly, at least qualitatively, for the radiolabelled L-chain.

In order to compare the efficiency of the degradation on such small amounts of material, the recovery of radioactivity was compared with the yields of PTH-amino acids from the carrier myoglobin as determined by GLC of the remainder of each fraction after conversion and extraction.

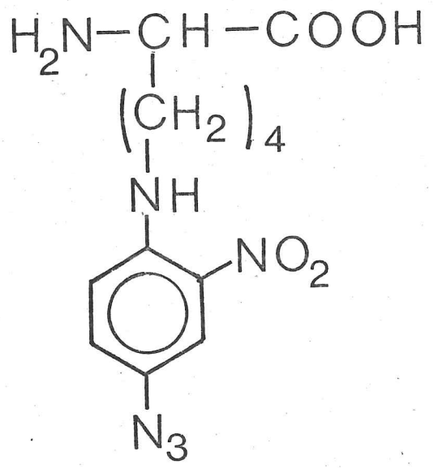


Fig. 3.11 ε-(4-azido-2-nitrophenyl-)lysine. Affinity label used in section 3.13.2.2 (Fisher & Press, 1974). The lysine moiety was ^{3}H-labelled (1.5×10^{7} - 2.5×10^{7} c.p.m./μmole).

and extraction.

The yield of ^{35}S-Met at step 4 (from Fig. 3.10) is 74.6%, allowing for sampling. The repetitive yield calculated from the yields at cycle 4 and cycle 11 is 89.3% and that from 4 and 13 is 88.9%. The yield of Leu (from the carrier myoglobin) at cycle 2 is 67.7 and the repetitive yield from cycles 2-9 is 88.9% and from 2-11 is 89.9%. The difference between the absolute yields is probably not significant as the amount of starting material for the carrier was not accurately measured. The similarity of repetitive yields suggests that the degradation reactions proceed equally efficiently on radiochemical amounts of sample. (The low repetitive yield of both labelled and carrier protein was due to a machine fault that was subsequently cured).

Using a similar approach Schechter (1973) has obtained some sequence information on the precursor to MOPC 321.

3.13.2.2 Rabbit H-chain - A second application of radioactive methods involved the use of a ^{3}H photo-affinity label, ε-(4-azido-2-nitrophenyl)-L-lysine (Fig. 3.11, Fleet et al, 1969).

A peptide from a rabbit antibody labelled with this reagent (the sample was given by Miss E. M. Press) was subjected to 30 cycles of degradation. The sample (360 nmoles) contained 332,000 cpm ^{14}C as S-carboxymethyl cysteine and 383,000 cpm ^{3}H as nitrophenyl-lysine derivatives of amino acids. Before beginning the degradation the protein film was dissolved in HFBA, washed with chlorobutane ("Wash 1", Fig. 3.12) and then subjected to a "dummy cycle" in which no PITC was added. The chlorobutane wash after the first cleavage step ("Wash 2") was also collected and tested for radioactivity.

After each cycle of Edman degradation, the thiazolinones were dried in the fraction collector as normal. After redissolving in 0.5 ml

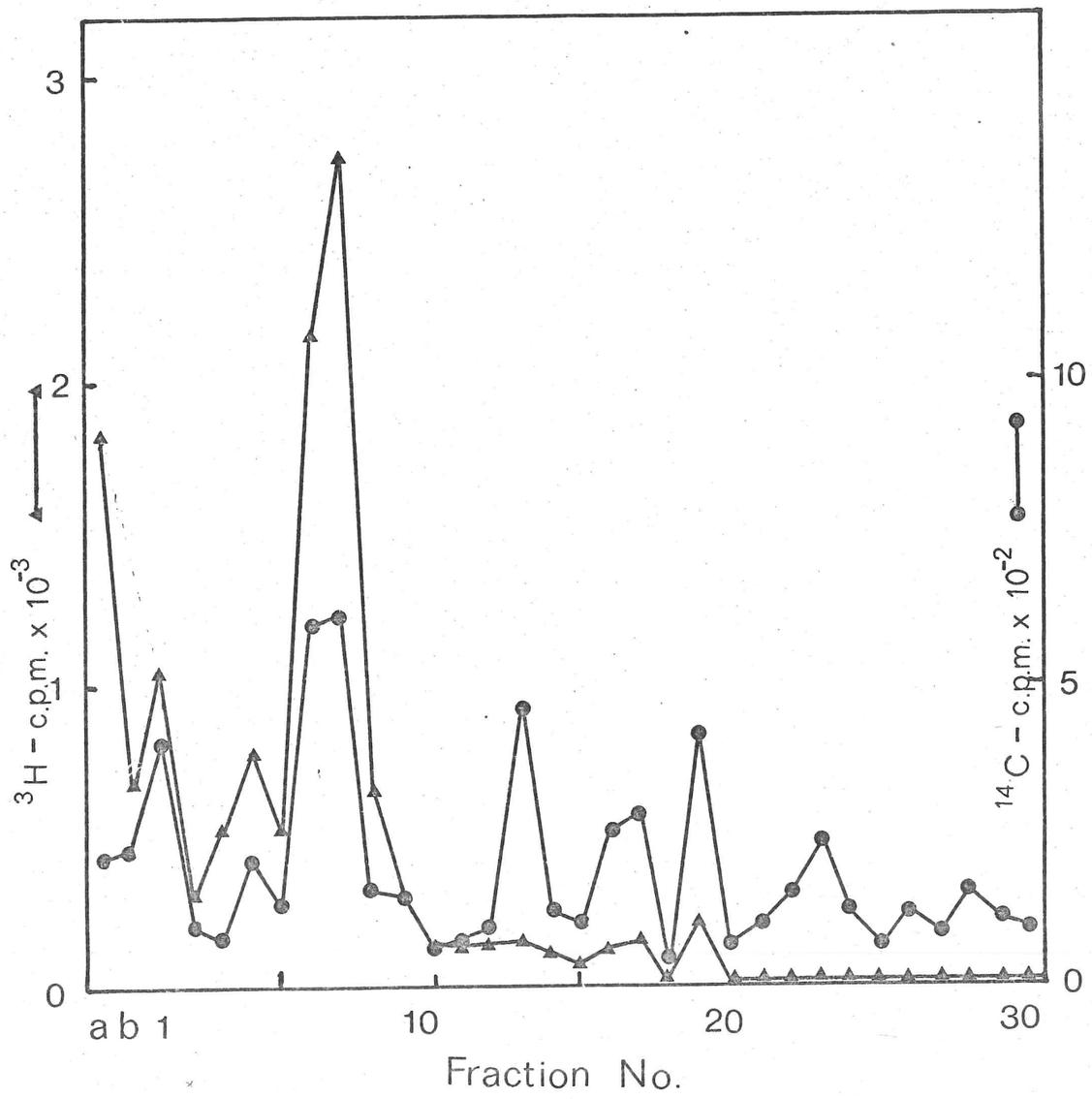


Fig. 3.12 Degradation of rabbit IgG peptide T3. The peptide was labelled with ^{14}C as ^{14}C-carboxymethyl cysteine and with the affinity label ^{3}H-ε-(4-azido-2-nitrophenyl-)lysine (Fig. 3.11). a and b ("Wash 1" and "Wash 2" in the text) were "dummy" cycles without R1 and show sample being washed out of the cup. The counts shown in every case are those measured in half the fraction.

TABLE 3.3

PTH-amino acids recovered at each cycle of Edman degradation of rabbit T3 peptide. Yield in nmoles (allowing for sampling) is in brackets where calculated.

Res. No.	Gas chromatography		Amino acid analysis	
	Major	Minor	Major	Minor
1	Ala (52), Gly (78)		Gly	Asp
2	Leu (54), Tyr (78)	Ser	Tyr (84), Pro, Leu (60)	
3	Leu (70)	Ser, Gly	Leu, Gly	
4	Pro		Pro	Gly, Tyr
5	Ala, Phe (13), Glu			
6	Pro			
7		Val		
8	Ala, Leu			
9				
10		Ser, Gly, Leu		
11				
12				
13				
14		Ala, Leu		
15				
16				
17		Val?		
18		Val?		
19				
20				

chlorobutane, 0.25 ml was removed for counting and the remainder dried, converted and extracted in the normal way. The PTH-amino acids were examined by GLC and amino acid analysis after HI hydrolysis.

Radioactivity was measured by liquid scintillation counting in a Nuclear Chicago Unilux II, using the dioxane base scintillant NE 250 (Nuclear Enterprises Ltd., Sighthill, Edinburgh, UK). The residue in the cup after 30 cycles was dissolved in 1.5ml 50% (v/v) pyridine and an aliquot counted.

The results of the analysis of radioactivity are summarised in Fig. 3.12. It can be seen that some ^{3}H was washed out of the cup before degradation was begun, that a larger amount was removed in the first 10 cycles and that very little was recovered in cycles 11-30. However the recovery of ^{3}H in cycles 1-10 amounts to only 5% of the total ^{3}H input. The ^{14}C is distributed more evenly amongst the first 30 residues, but with a peak at residue 6 amounting to 0.65% of the possible ^{14}C yield. GLC and amino acid analysis showed a high and variable background of PTH-amino acids characteristic of a heterogeneous sequence. It was possible however to identify some PTH-amino acids as major components of the first few cycles (Table 3.3). Positions 1 and 2 contain at least 2 major components which account for most of the expected yield. There are also minor components. Position 3 contains a major residue; positions 4-6 were difficult to quantitate owing to the nature of the amino acid derivatives present; and thereafter no major components were detected.

The material remaining in the cup after 30 cycles contained 13% of the input ^{3}H radioactivity and 18% of the ^{14}C.

Fig. 3.12 suggests that the affinity label may be bound predominantly to residues 6 and 7 and almost exclusively within the first 10 residues. The reason for the very low recoveries was not discovered. It might

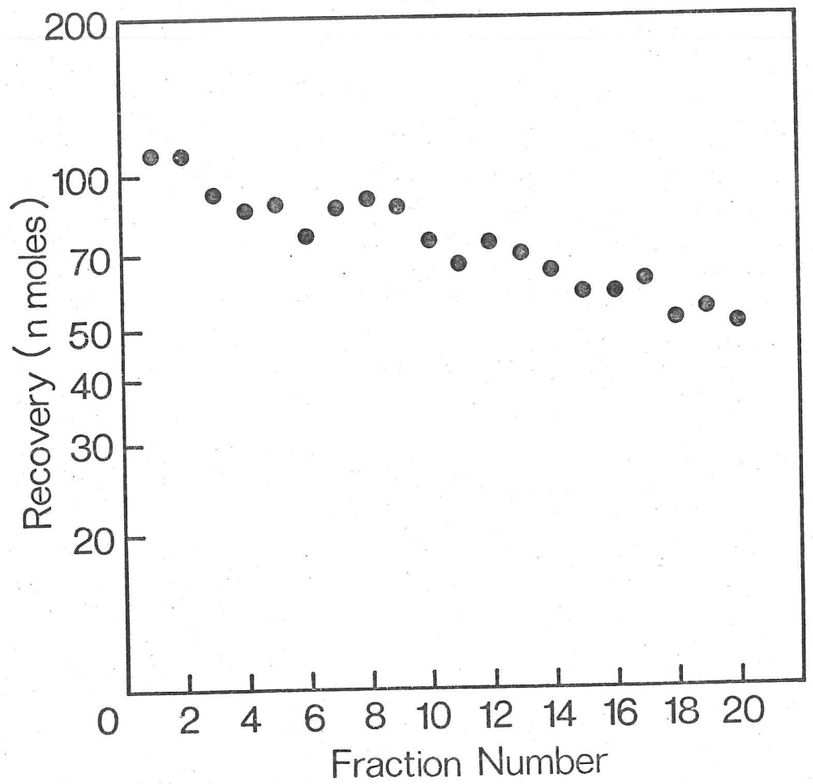


Fig. 3.13 Automatic degradation of alkaline phosphatase. Corrected semi-log plot of amino acid recoveries against fraction number. Yields of PTH-Ala, -Gly,-Val,-Leu,-Ile,-Met were calculated from GLC peak heights. Arg, Pro, Glu, Asp yields were calculated from amino acid analyses.

have been due to steric hindrance of the affinity label (Fisher & Press, 1974) or to reaction of the highly reactive nitrene with the peptide bond and consequent interference with the Edman degradation.

3.13.3 Mixture analysis

In the absence of comprehensive accurate methods for the estimation of PTH-amino acids, the complex mixture analysis envisaged by Gray (1968) is not yet feasible. However simple mixture analysis was shown to be possible.

The first 8 residues of IF-1 H-chain were determined from an equimolar mixture of H- and L-chains prepared from the myeloma protein IF-1. The sequence of the L-chain was known and the H-chain sequence could thus be deduced by subtraction (Chapter 6).

Proof of the absence of heterogeneity (at about the 10% level or above) can be convincingly shown. The N-terminal sequence of E.coli alkaline phosphatase (E.C.3.1.3.1) was investigated using the sequenator. Previous reports had suggested N-terminal heterogeneity as the basis for isozymes of this enzyme (Natori & Garen, 1970). Degradation of 200 nmoles for 20 cycles revealed the unique and unambiguous sequence:

*Arg-Thr-Pro-Glu-Met-Pro-Val-Leu-Glu-Asn-Arg-Ala-Ala-Glu-Gly-Asp-Ile-Thr-Ala-Pro-. The yield of the principal amino acid at each step is shown in Fig. 3.13. The repetitive yield was 97% and at cycle 20 overlap of the previous residue was 14% and the "signal to noise" ratio 6:1.

---

*Contamination of the protein preparation with free amino acids resulted in apparent heterogeneity for one residue. This was clarified by dansylation without hydrolysis (Bridgen and Secher, 1973).

CHAPTER 4

N-terminal Sequences of Rat Myeloma L-chains

CHAPTER 4

N-terminal sequences of Rat Myeloma L-chains

4.1 Introduction

The previous chapter described a technique of automatic sequence determination with which a homology search of the partial (N-terminal) sequences of many proteins can be made. This technique has been used to perform a survey of 25 rat myeloma L-chain N-terminal sequences. In this chapter the results obtained are described and discussed with reference to two questions:

1) How does the pattern of variability compare with the equivalent data from human immunoglobulins and with that from the much more closely related Balb/c strain of mice?

2) What are the implications for the development of immunoglobulin genes, in evolution and during the somatic development of the individual?

4.2 Reconstruction of protein evolution

As early as 1958 Crick argued that "amino acid sequences of the proteins of an organism ....are the most delicate expression possible of the phenotype of an organism and that vast amounts of evolutionary information may be hidden away within them" (Crick, 1958).

In 1963 a "method of minimum evolution" was stated (Edwards & Cavalli-Sforza, 1963) and used as the basis for construction of statistical methods for the comparison of blood group gene frequencies (Edwards & Cavalli-Sforza, 1964) and, in a different form, for a comparison of cytochrome C from different species (Fitch & Margoliash, 1967). Similar procedures have now been used to compare proteins

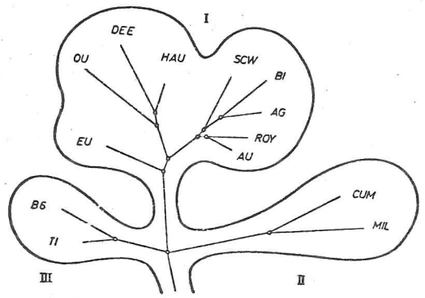


Fig. 4.1 Genealogic tree for the evolution of Human κ chains whose complete sequences are known (Eulitz and Hilschmann, 1974). The V-regions are grouped according to sequence homology using a modified method of Dayhoff (1972). The branchpoints represent hypothetical ancestral sequences and the length of the branches represents the difference between a sequence and the ancestral sequence. I, II, III refer to the κ-sub-groups.

whose sequence is known in many species (see Dayhoff, 1972) and, in general, the phylogenetic trees which are deduced are in agreement with the independant fossil evidence. Thus, the earlier two species diverged in evolutionary time, the more different will be their protein sequences due to the random and constant accumulation of mutations. For Ig's the various myeloma proteins from a single species may be compared in a similar way and their "evolution" reconstructed. How does such an analysis distinguish between germ-line evolution and somatic evolution, which takes place over again in each generation? i.e. What level of a genealogic tree represents the germ-line genes?

Where the difference between two related proteins is large, if they are both present in all individuals of the species then a large number of identical somatic mutational events would have to take place independently and in the same order in each individual. Data on somatic mutation are very scant but our understanding of mutation in prokaryotes suggests that this is highly unlikely.

Thus evidence of the above type is accepted as proof that, for example, κ and λ C-genes arose from a common ancestral C-region gene by a gene duplication about 400 x 10^6 years ago (see Fig. 4.1). However where the difference between two sequences would require only a few mutational events the consequences of somatic and germ-line evolution may be difficult to distinguish.

If this analysis is extended to the V-region sequences of, say, human κ-chains, an evolutionary tree can be constructed (Fig. 4.1). The subgroups (chapter 1) are clearly separated, but within each subgroup the differences between the sequences are so small that it is impossible to describe confidently a uniquely best-fitting tree since the choice between alternative genealogies might be confused by a single

back mutation or parallel mutation. (For a discussion of these problems see Smith, 1973). Even more important perhaps is the possibility that the substitution of an amino acid by mutation at a particular site may not be a random, independent event, but at least at some sites, may be highly dependent on the sequence and three-dimensional structure of the protein (Milstein & Svasti, 1971 Novotny, 1973) or DNA (Koch, 1971).

Proponents of germ-line theories place great stress on the nature of replacements between two myeloma proteins of the same subgroup (Hood, 1973; Hilschmann et al, 1969). The linked interchange of two pairs of amino acids is presumed to indicate at least two structural genes. Perhaps examination of atomic models of Ig's will suggest that some of these linked interchanges are dictated by structural requirements.

To account for the fact that all human sequences fall into subgroups a germ-line theory postulates a large recent expansion of the three major subgroup genes. Distinct clusters of similar sequences are, conversely, exactly the pattern expected from somatic diversification proceeding independently in each Ig-producing or precursor cell.

The genealogic tree derived from Balb/c κ-chain N-terminal sequences is much less well defined. The boundaries of subgroups are unclear, at least from the sequences available so far, and a more general term such as "sets" is preferable to "subgroups" which, at least in its original usage, has a more precise genetic implication.

4.3 Materials and Methods

The L-chains were a gift from Dr P. Querinjean and were prepared from myeloma proteins isolated from rats of the LOU/Ws1 strain (Querinjean et al, 1972). About 5-20 mg (usually 15 mg) of reduced

and ^{14}C-carboxymethylated protein was used for a single run. About half of the runs were repeated to confirm tentative assignments or to benefit from technical improvements. The samples were usually dissolved in distilled water or, where necessary in very dilute ammonia solution (obtained by blowing ammonia vapour onto the surface of the liquid). A few proteins were insoluble in this too and these were applied as a suspension (with no adverse effects on the success of the run). The operation of the machine was as described in the previous chapter except for the "short single cleavage" (see below).

4.4 Results

4.4.1 Improvement of repetitive yield

All the proteins were found to have an unblocked N-terminus and were thus amenable to the Edman degradation.

A problem encountered in some early runs was the apparent blockage of the protein after five cycles of degradation. Residue 6 is an invariant Gln and the blockage was presumably due to the cyclisation of the newly exposed N-terminal Gln to PCA. This same problem was noted in the comparison of mouse L-chains undertaken by Hood *et al* (1973). The reaction is acid catalysed and the problem was overcome by modifying the program during the fifth cycle to reduce the exposure of the Gln to acid conditions before the next coupling reaction. Thus in place of the standard double cleavage with HFBA, a single cleavage reaction was substituted by manually overriding the program and the reaction time reduced from 120s to 10s. There was no indication of increased overlap or decreased yields following this short single cleavage. Presumably the bond between the derivatised Thr and the Gln at position 6 is very labile to acid cleavage so the procedure was adopted routinely in the 5th cycle, without first establishing whether or not it was required for a particular protein.

```
                5        10        15        20        25        30a b c d e f 31      35        40
IR97    D I V L T Q S P - V L A V S L G Q R A T I S C R A(S)Q(S)V(S)I S ? I H L M H - - W Y Z Z K P G(E)
IR9     D I V L T Q S P - V L A V S L G Q R A T I(S)C R A(S)Q S V(S)T ? ? Y B L M H - - W Y Z(Z)K P G Z Z Z
IR11    D T V L T Q S P - A L A V S P G E R V S I S C R A S Z S(V)(S)? L ? ? - - - - - - W Y

IR25    D I Q M K Q S P A S L S A S L G E T I S I E C L A(S)Z B I ? ? - - - - - - Y L A ? Y
IR15    D I Q M T Q S P A F L S A S L G E T V S I E C L A S ? D I Y ? - - - - - - - Y L W A F ? Q
IR12    D I Q M T Q S P A S L S A S L G E G I S I E C L(K)S Q

S208    D I Z M T Z S P S L L S A S V G B R V T L S C K A G Q K I B G - - - - - - Y L
S216    D I Q M T Q S P S L L S A S V G D(R)V T L B C K A S Q ? I B K B - - - - - - L E W - E
S221    D I Q M T Q S P S L L S A S V G D R V T L N C K A S Q ? I Y Y ? - - - - - - L A W
IR52    D I Q L T Q S P S L L S A(S)V G D R V T L S C K A(S)Z(B)I(Y)
IR162   S I Q V T Q S P S L L A A S V G D S V T L S C K A

IR62    D V Q M T Q S P S Y L A A S P G E S V ? I S C K A ? ? ? I
S211    D V Q M T Q S P S Y L A A S P G Z S V S I S C K A S N K S I S N N - - - - - L A W Y Q Q K

IR32    D I V M T Q S P T S M S I S V G D R V T M N C K A S Z B(V)G ? S V G - - - - - - W ? ? Y
S223    D I V M T Q S P T S M S I S V G D R V T M N C K A(S)Q N V G

S210    D F V M T Q S P S S L A V S A G E T V T I N C K S S Q S L F Y S G N Q K B Y L A W Y Q Q K P G E ? P K L
IR27    D I V M T Q S P S S L A V S A G E T V T I N C K S S Q(S)(V)L
S204    D I V M T Q T P S S Q A V S A G E K V T M N C R S S Q(S)L L Y S Z B K K B Y L A ? Y Z Z K P G Z H

IR55  G Z I V L T Q A P L G V(V)?(T)P G E Z A ? I S C R Y G Z K L Y
IR74  G Z I V L T Q A P L S V ? ? G P G E(E)A(S)I(S)C

S207    D I Q M T Q S P A S L S A S L E E I V T I K C Z A Z K K I G ? - - - - - - Y L A ? Y

IR33    D I Q M T Q S P P S L(S)V(S)L G D(K)V T I T C Q ? ? Q(N)I N ? ? - - - - - - L A ? Y Q

S220    D V V L T Q T P S I L S A T I C Q S V S I S C

IR24    D V L L T Q T P T ? L P V S L G G Q V S I S C R S ? Q ? L V

S218    D I M L T Q S P A T L S V T(P)G E S V S L S C R A S Q(R)I G
```

Fig. 4.2 N-terminal sequence of 25 rat κ-chains determined using the sequenator. The sequences are grouped in "sets" according to homology (see text and Fig. 4.4). ( ) = provisional assignment; - = assumed deletion; ? = unidentified residue. IR55 and IR74 were heterogeneous in length (see text) but only the longer sequence is shown here. All sequences were determined in the course of this study except IR162 (Querinjean & Capra, personal communication).

4.4.2 Presentation of sequence data

The results of the sequenator runs are summarised in Fig. 4.2. The one-letter code is used to indicate the major ("sequence") residue at each position. In the case of three proteins (IR97, IR9, IR11) a deletion of one residue has been assumed in order to maximise homology with the other proteins. Otherwise each sequence was independently determined without reference to the others. In general, for the first 30 residues two methods of identification were required for an unambiguous identification of a sequence residue. Where only one procedure (GLC, TLC, or amino acid analysis) gave an indication of the residue at that position, or where the assignment was for any other reason tentative that residue is shown in parenthesis. After position 30 no such distinction was made and all assignments should therefore be regarded as tentative. Positions at which no clear indication of the sequence residue was obtained, are left blank.

It can be seen that there was considerable variation from protein to protein in the degree of success of the sequence determination. With some proteins the majority of residues were unambiguously assigned for 45-50 cycles, whereas in other cases problems in residue identification occurred within the first 25-30 cycles. Possible causes for this wide variation are discussed below.

4.4.3 N-terminal heterogeneity

The sequence analysis of proteins IR55 and IR74 was more complicated than the others. The N-terminal residue appeared heterogeneous (mainly Gly and Glx) and there was about 100% overlap at each cycle. The level of double overlap (i.e. the residue two previous) was normal. The same results were obtained on repeat runs. It was possible to interpret the data as two similar sequences, differing only by the addition of Gly to the N-terminus in one case and only the longer

Species	No. of proteins	Position 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Human	45	D34 E10	I43	Q27 V18	M31 L14	T45	Q45	S44	P45	S26 G8 L7 A4	S28 T16	L45	S38 P7	A23 L10 V9 T2	S36 T7 V2	V26 P18	G44	D30 E15	R38 P7	V26 A18	T34 S8	I33 L12	T24 S18 A2
		K M V						T			P			M		L	R			I	A		Z
Balb/c Mouse	44	D34 E8 N2	I32 V8 T2 N2	V31 Q9 T2 L2	M26 L14 T2 V2	T42 I2	Q44	S34 T8 B2	P39 T2 E2	A20 S10 L8 T2 K8 B2	S27 F6 T5 I4	L31 M9 D2 T2	S24 A11 P4	V26 A12 M3 T3	S33 T9 A2	L16 V10 I6 A6 P3 S2	G42 S2	E18 D8 Z6 Q5 K3	R17 K8 S4 T2 E2 P2	V32 A12	T26 S14	I29 L5 M5	S21 T15 []2
											L Y		F L T V Q			T		L B	L B			V	
LOU/Ws1 Rat	25	D22 Z2	I19 V4	Q12 V11	M15 L9	T24	Q25	S20 T3 A2	P25	S11 A5 T3 L2	S10 L5 Y2	L20 M2 V2	S13 A9	A12 V9 I2	S21 T3	L8 V7 P6 A3	G24	E12 D7 Q3	R9 S5 T4 K2	V19 A4 I2	T14 S9	I16 L6 M3	S13 N6 E3
		S	T F	L M	V	K				P	A F I T G	Q	V P		G	I	E	B G Z	G E Q Z I				T K B

Fig. 4.3 Diversity in N-terminal 22 residues of human, mouse and rat κ chains. The alternative amino acids observed at each position are shown, together with their frequency (except where only occurred once). Data from fig. 4.2 and Hood et al. (1973). [] = deletion.

sequence is shown in Fig.4.2. The possible cause and significance of this observation are discussed below.

4.4.4 Sequence variation and comparison with other species

A visual comparison of the various sequences shows a clear homology ranging from 100% (IR97 and IR9) to 33% (e.g. IR55 against S207) over the first 30 residues. The variation between proteins is not evenly distributed over all 30 positions. Residue 7 for example is always Pro, residue 6 Gln (allowing one Glx as Gln) and residue 23 always Cys. Some positions are almost invariant (1,5, 7,12,14,16,19,26,27) and others have two or more common alternatives. This data is summarised in Fig.4.3 which gives the alternative residues found in each of the first 22 positions and the frequency of that residue. For comparison the corresponding data for 44 mouse L chains and 45 human chains are also given (Hood et al, 1973).

These authors noted that the mouse chains exhibited a greater variation than the human sequences. The rat proteins are clearly even more diverse as a similar degree of diversity has been found in about half the number of proteins examined in the other species.

The positions of restricted variability show a strong homology between the three species; the somewhat greater homology between rat and mouse than between either rodent and man is consistent with the more recent divergence of the rodent species.

4.4.5 Computer comparison of sequence data

A computer comparison of the N-terminal sequences of the 25 rat L-chains was carried out using the program described by Milstein and Pink (1970). The program, which is essentially that of Fitch (1966) compares each sequence in turn with all the others by selecting every possible pair of sequences and computes the minimum number of single-

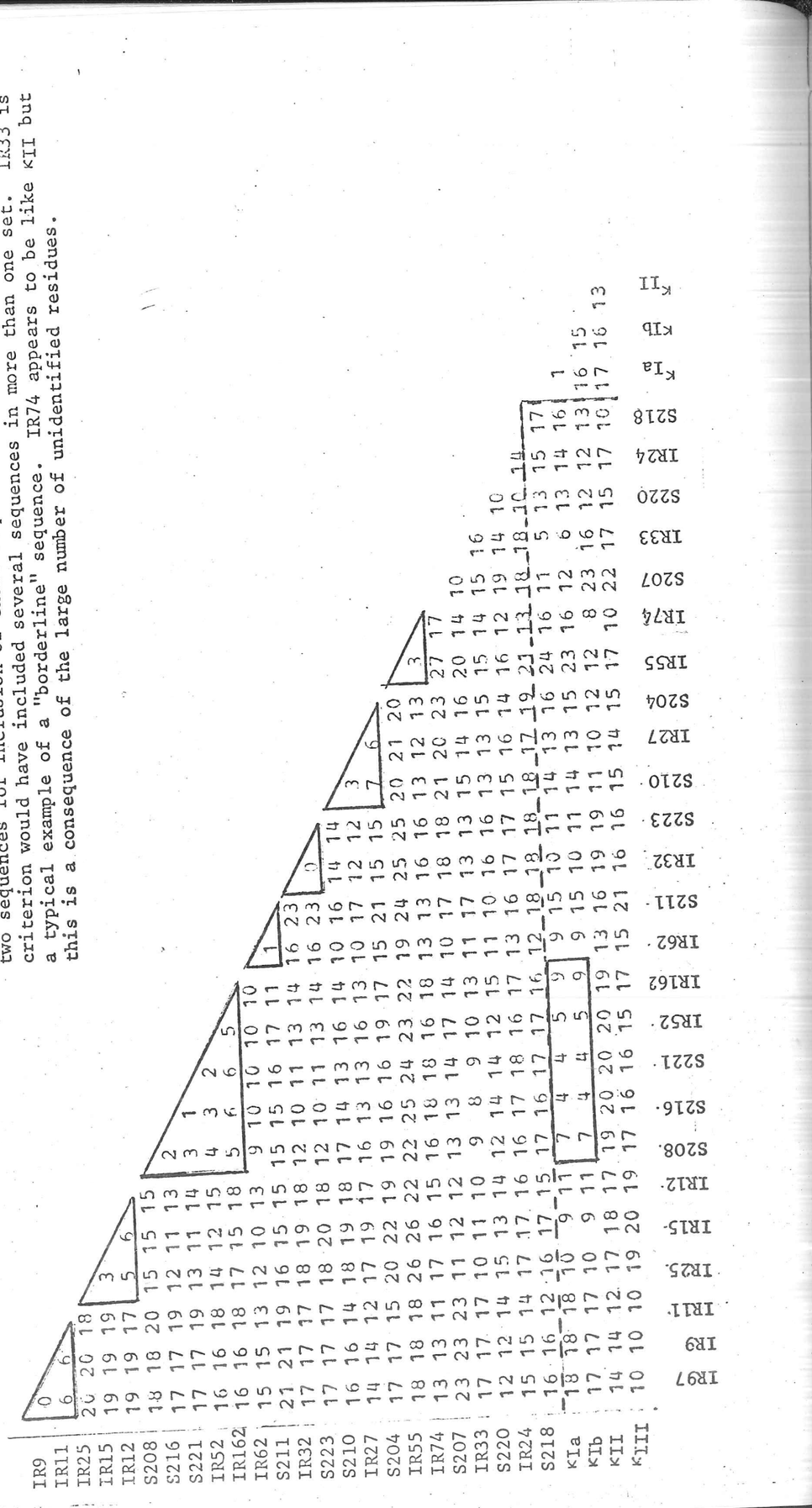


Fig. 4.4 Computer comparison of rat κ-chain N-terminal (30 residues) sequences. The numbers in the matrix indicate the minimum number of base changes that are required to convert one sequence into another. The programme is essentially that of Fitch (1966) as described by Milstein & Pink (1970). Black lines enclose the comparisons of sequences within a "set" (see text). Broken lines mark the boundaries of the rat sequences from the human V_κ sequences which were also compared. 7 base changes per 30 amino acids was chosen as the limit between any two sequences for inclusion of those sequences in a set since a less stringent criterion would have included several sequences in more than one set. IR33 is a typical example of a "borderline" sequence. IR74 appears to be like κII but this is a consequence of the large number of unidentified residues.

step mutations necessary to derive one sequence from any other. No attempt is made to calculate most probable genealogies. Certain simplifications to the data of Fig. 4.2 were made for these calculations:

1) Only the first 30 residues were included.

2) Tentative assignments were given the same weight as firm ones.

3) A deletion was counted as 2 nucleotide differences from any residue.

4) No extra differences were calculated for the apparent N-terminal addition of IR55 and IR74.

5) An unidentified residue (U) was allowed identity with any other residues.

6) Glx was allowed identity with Glu or Gln; similarly Asx with Asp or Asn.

After a few preliminary runs to group together those sequences of high homology the data shown in Fig. 4.4 was obtained. These show that some of the sequences may be grouped together into sets such that a sequence within a set resembles any other sequence within the same set more closely than it resembles any sequence outside that set. However in contrast to the equivalent comparison of human sequences the sets are many and small, the boundaries between sets are ill-defined and some sequences, which may represent further sets, do not fit into any particular set.

Included in the comparison of Fig. 4.4 are the four human subgroup sequences. Comparison of the individual rat sequences with the human prototype sequences shows that the set containing S208, S216, S221, IR52 and IR162 resembles the κ_I sequences more closely than the other human basic sequences. There is no close relationship between any of the rat proteins with the κ_{II} or κ_{III} sequences.

4.5 Discussion

4.5.1 Methodology

The data collected on the rat L-chain sequences illustrate one of the most powerful and labour saving applications of automatic sequence determination to date. The comparative data on human L-chains had to be collected from the complete sequences determined manually by several groups over a period of many years. These results showed that the N-terminal sequences were representative of the whole proteins and that a comparison of N-terminal sequences was therefore a valid approximation to total sequence comparison. Using the sequenator this approach had been directly applied to a survey of mouse L-chains (Hood et al, 1973) and now the corresponding data from 25 rat proteins has been collected in only 18 months. As a result of technical developments since then this time could now be reduced to less than half.

During the work on these and other proteins it became apparent that different classes of proteins differ in their amenability to sequenator analysis. Sometimes the reason for this was obvious (e.g. the rearrangement of an Asn-Gly-) or suspected (the presence of large amounts of carbohydrate) but in other proteins it was presumably due to some undefined aspect of the particular sequence or type of sequence. A poor run may reflect for example unusual lability of the protein to non-specific cleavage resulting in a high background or a high proportion of amino acid residues which react with PITC or cleave in HFBA less favourably. In this respect the rat L-chains were certainly non-ideal proteins. The average length of run was considerably less than in the case of other proteins run under similar conditions. Nevertheless it was possible with most L-chains to define a sequence up to 30 residues, compared to the 23 residues achieved with the mouse L-chains. Careful and complete quantitation of yields could have provided a few more residues in most runs, but since 31 represents

	31			35						40					45		
Human	Leu	Asn Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Lys Gln	Ala Ser	Pro	Lys Glx Arg	Leu	Leu	Ile	Tyr
Mouse	Met	Gln	Trp	Tyr	Glx	Glx	Lys	Pro	Gly	Glx	Pro	Pro	Lys	Leu	Leu	Ile	Tyr
Rabbit	Leu	Ala	Trp	Tyr	Glx	Glx	Lys	Pro	Gly	Gln	Pro	Pro	Lys	Leu	Leu	Ile	Tyr
Rat	Leu	Ala	Trp	Tyr	Glx	Glx	Lys	Pro	Gly	Glx	?	Pro	Lys	Leu	?	?	?

Fig. 4.5. Comparison of the predominant sequences following the first hypervariable region of κ-chains. For humans the alternative residues in the three sub-group basic sequences are shown. Data from Milstein & Pink (1970), Smith (1973) and fig. 4.2.

the beginning of the first hypervariable region, 30 residues is an obvious choice for N-terminal comparisons.

The possibility that rat myeloma proteins are apparently difficult to sequence automatically because of contamination with normal IgG L-chains (which might not be apparent in electrophoretic or N-terminal tests of homogeneity) has not been excluded. Another reason for apparent problems in achieving long runs was the occurrence of sequences containing a high proportion of Ser or other residues difficult to identify. Some of the rat L-chains contain the sequence Ser-CMCys-X-Ala-Ser-X-Ser and this was very difficult to assign unambiguously.

However the L-chains were not uniform in this respect. The variation in the relative success of sequenator runs on different proteins was striking. Although in part no doubt due to variation in protein purity and in performance of the equipment (sequenator and gas chromatograph columns) repeated runs suggested that the ease of sequencing a given protein was a characteristic of that protein.

The blocking of residue 6 is a particular example of the variation between proteins. The extent of blocking in early runs, before the introduction of the short single cleavage ranged from almost total obstruction of the run to undetectable blocking, despite the highly conserved sequence (-Thr-Gln-Ser-Pro-) around residue 6 (Fig. 4.2).

4.5.2 Strong homology following first hypervariable region

The 4 longest sequences possess almost complete identity in those residues following the first hypervariable region although they do not all have similar sequences in the first 30 residues, nor even similar sequences or the same number of residues in the hypervariable region. This suggests a high conservation of sequence in this region. Fig. 4.5 shows a comparison of residues 31-45 of the rat chains with the homologous residues in mouse, rabbit and human κ chains. The high degree of homology

suggests an important functional or structural role for this region of the molecule.

4.5.3 Deletions and additions

The existence of deletions in the sequences IR97, IR9 and IR11 was inferred from the unusual position of the radioactive CMCys (cycle 22) and from the overall homology. The assignment of the deletion to position 9 was to maximise homology between these three sequences and the other 22. It could be that in the case of IR11 the deletion should be at position 10 as there is more homology for Ala at position 9 than 10. However this change makes little difference to the overall grouping or to the fact that a deletion at position 9 or 10 has not been observed previously in the L-chains of other species.

The detection of N-terminal Gly (IR55 and IR74) is also unique to the rat κ-chains. Its significance and also that of the N-terminal heterogeneity is not understood. The following possibilities were considered:

1) That two V-regions are synthesised by a single tumour. It is surprising that the only sequence difference should be an addition at the N-terminus and that it should be the same in both cases.

2) That the heterogeneity reflects a frequent incorrect initiation in the translation of the Ig mRNA. This would be the first example of such mis-translation.

3) If Ig L-chains are synthesised *in vivo* by proteolytic cleavage of a precursor containing additional residues at the N-terminus (Milstein *et al*, 1972), it could be that the sequences of IR74 and IR55 result from an incorrect cleavage of the precursor. This could account for the "addition" of the same amino acid in both myelomas. Whether the false cleavage might be the direct consequence of the precursor of Glx rather than the normal Asp, or whether the Glx is merely indicative

of the set of proteins with the required properties, cannot be deduced from the data so far. The presence of Gln at the N-terminus of several human and mice proteins shows that in these species at least the N-terminal Gln does not entail imperfect cleavage.

4.5.4 Non-random variation

The distribution of variability within the first 30 residues was similar to that in humans and mice. The highly restricted positions (2,5,6,8,16,23) are also highly homologous and presumably reflect rigid structural requirement common to the L-chains of all 3 species. Fig. 4.3 showed that the extent of variation is much greater than that observed in humans or even in mice. This suggests that, in the pool of myeloma proteins studied, the rats are expressing a greater number of V-genes. What indications do the sequences provide as to whether this variety arises by virtue of a larger germ line V-gene pool or by more extensive somatic diversification?

4.5.5 Species-specific residues

Doolittle suggested that the existence of residues which occurred at a given position with a high frequency in one species and were absent or at a low frequency at the corresponding position in another species implied that at the time of divergence of the two species there was only a single gene coding for those V-regions characterised by the common ancestral species-specific residue. Some time after divergence of the two species the single gene mutated at that position in one of the species. Variants of these sequences must have arisen subsequent to this mutation (and hence subsequent to the divergence of the species) either by massive gene duplication and further mutation or by somatic diversification.

The rat L-chain data now allows comparison between two very closely related species whose evolutionary divergence is estimated at less than 25 million years ago (Britten & Kohne, 1967). There is no outstanding example of a species specific residue between these species. (The apparent candidate at position 21 results from a typrographic error in Fig. 6 of Hood et al, 1973). Asn occurs at position 22 in 7 of rat L-chains examined and in none of the human or mouse proteins. Leu occurs at position 10 in 5 of the rat proteins but not in the humans and in only one out of 44 mouse chains. However these examples occur in a minority of the rat L-chains and it seems equally likely that they represent genes which could have duplicated and mutated before species divergence and have been lost, or not yet detected in mice.

4.5.6 "Sub-groups"

The definition of sub-groups in humans (Milstein, 1967) led to the suggestion that there are 3 V_λ genes in the human germ-line and that further diversification arises by somatic mutation of these genes. In humans the definition of sub-groups was clear because of the existence of few sub-groups with relatively large differences between their prototype sequences.

The computer comparison of the rat L-chain sequences clearly shows that they do not fall into such a simple sub-group pattern. If the sets of similar sequences represent the same phenomenon as the sub-groups in humans then there is a much higher minimum number of V_κ germ line genes in rats than in humans and this number is too large to define from the sample already studied. The difficulty of sub-group definition is reflected in Fig.4.4 by the borderline sequences e.g. IR33 and IR62 which show a slightly greater similarity with the sequences within the set S208, S216, S221, IR52, IR162 but which cannot be included by any overall definition of the maximum difference permitted between any two members

of a sub-group.

It may be that the sets of rodent sequences are not completely analogous to the human sub-groups. Smaller animals with a shorter gestation time may not have sufficient time to develop a sufficiently wide repertoire of V-genes by birth and may therefore require additional mechanisms for generating diversity. It is also of interest that the large number of V_κ-genes indicated from the sequence data of rats and mice correlates with the high proportion (>95%, Hood et al, 1967) of κ chains in the pooled sera of normal animals.

It now seems most unlikely that the complex pattern of mouse κ chain sequences results from the unusual origin of the tumours (e.g. mineral oil injection), since the rat tumours used in these studies all arose spontaneously and yet the pattern of diversity is similar to that seen in the mouse study.

Simple recombination has been ruled out as a mechanism for generations V_κ diversity from a few germ line genes in humans (Milstein et al, 1969b; Smith et al, 1971). However multiple recombination between several germ line genes as suggested by a recent model of Smithies (1970) for example, with further elaboration by somatic point mutation could provide the complicated comparisons observed in Fig. 4.4.

4.5.7 Do different species have common V-genes?

Fig. 4.4 also contains a comparison of the human prototype sequences with the individual rat sequences. This shows a marked homology between IR33, S 208, S221, S211, IR52, IR62 and the human κ_{Ia} and κ_{Ib} sequences. No other rat sequences showed a similar degree of homology with the κ_I sequences, nor to any other human basic sequence. It is interesting that it is only to the κ_I sequences that some mouse κ-chain sequences appear to be related. This suggests that the genes maintained in each species may not have been chosen at random from an

ancestral germ-line pool, but that the κ_I sequence may have a selective advantage in all three species (Milstein & Svasti, 1971).

CHAPTER 5

Clonal Variation of P3K cells in Tissue Culture

CHAPTER 5

Clonal Variation of P3K Cells in Tissue Culture

5.1 Introduction

This chapter describes the development of a technique for the screening of a population of cells for spontaneously arising mutant clones and its successful application resulting in the isolation of two such clones. The cell line studied (P3K) derives from the mouse plasma cell tumour MOPC21 and was adapted to tissue culture by Horibata and Harris (Horibata & Harris, 1970). The cells secrete a single major protein, an Ig which appears to confer no benefit on the cell that produces it. In this situation cells producing altered Ig should, in general, not be selected against. Thus many mutations should be tolerated, giving rise to clones of mutant cells within the population. Furthermore it can be inferred, from the variety of sequences which exist, that the Ig molecule can "accommodate" many amino acid substitutions and still be recognised by the machinery for exporting the molecule from the cell. A third consideration in the choice of the P3K line was the extensive information available on protein sequence (Svasti & Milstein, 1972a,b) and mRNA preparation (Brownlee et al, 1973 , Cowan & Milstein, 1973).

Effective screening of cell populations is generally accomplished by the use of selective techniques in which only variant individuals survive. In the case of a myeloma cell culture such a selection could be based on the antigenic properties of the secreted Ig, or, for a myeloma protein of known antibody activity, on a change in this

activity. The former approach has been employed in studies of myeloma cell cultures (Coffino et al, 1972) and has revealed the spontaneous occurrence of cells which cease to secrete Ig H- and (subsequently or simultanteously) L-chain. However, selective techniques can detect alterations in only a small portion of the Ig and would be unable to provide information on the relative mutation rates in different parts of the molecule. Therefore a conceptually simpler approach was adapted in which cells taken at random from a continuous culture are examined for spontaneous variation in the Ig they secrete.

5.2 Isoelectric Focusing

Isoelectric focusing on polyacrylamide gel slabs was chosen for examination of secreted Ig (Awdeh et al, 1968). This technique combines high resolution and (by using radioactively labelled proteins and autoradiographic identification) high sensitivity. Since proteins in the sample come to equilibrium at their isoelectric point in a pH gradient of polyelectrolytes, no special care in the application of samples is required and this is a major consideration when a large number of samples are examined. Preliminary experiments with variants of haemoglobin (Ingram, 1957) and the B subunit of tryptophan synthetase (Cotton & Crawford, 1972) which in both cases differ from the wild-type protein by a single charge change showed that (in these cases at least) the method is capable of detecting such changes. Derivatives of bovine plasma albumin (Spencer & King, 1971) and of Ig L-chains (Feinstein, A., personal communication) differing only by a single charged group are also resolved by this method.

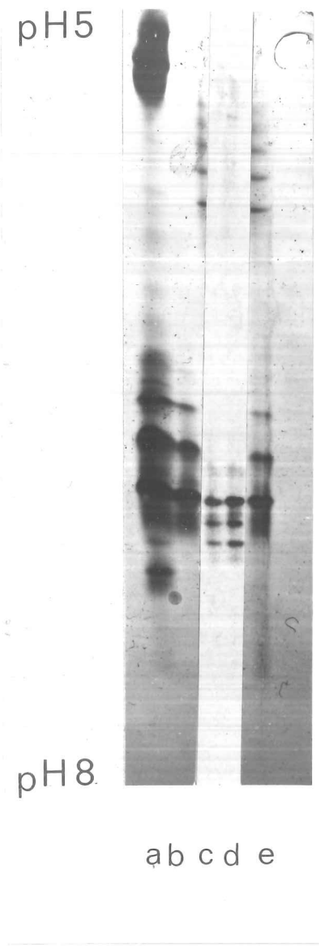


Fig. 5.1 Isoelectric focusing gel (pH 5-8). Composite photograph of stained gel (a,b,e) and autoradiograph of the same gel (c,d). b and c are in fact adjacent samples as shown. d and e are the same sample. a) Serum from mice bearing P3K tumour; b) purified P3K myeloma protein; c) ^{14}C-Lys labelled P3K IgG synthesised by cells in culture (see section 5.2.1); d) and e) mixture of samples b) and c).

5.3 Preliminary Experiments

Examination of the serum of tumour-bearing mice or of purified PK3 myeloma protein, revealed a series of well-defined protein bands in the region pH 6.5-7.5 (Fig.5.1) . When P3K cells were incubated for 24 hr. in the presence of ^{14}C-Lys and the extra-cellular medium analysed directly by isoelectric focusing a similar series of three major radioactive bands was obtained on radioautography. These bands were of higher isoelectric point than those obtained from unlabelled serum protein, but incubation of the radioactive extracellular medium with an equal volume of normal mouse serum for 16 hr. at 37^{o} converted the radioactive bands to the more acidic pattern. In the absence of added mouse serum only a slight degree of conversion was obtained, due perhaps to the dilute horse serum present in the medium. Rabbit serum, under the same conditions, was only about half as efficient as the homologous serum. Kölsch obtained similar results to these and showed that the serum factor responsible for the conversion is inactivated on heating to 60^{o} (Kölsch, 1967). These results are in accord with those of Awdeh et al who suggested that this conversion is a process of deamidation (Awdeh et al, 1970).

When extracellular medium was analysed after incubations of P3K cells with ^{14}C-Lys for 2, 24, 48, 72 hr. it was found that the optimal time for incorporation was 24 hr.

5.4 Examination of cloned cells

The secreted Ig of P3K cells was also examined after cloning the cells in soft agar (Cotton et al, 1973). ^{14}C-Lys was incorporated

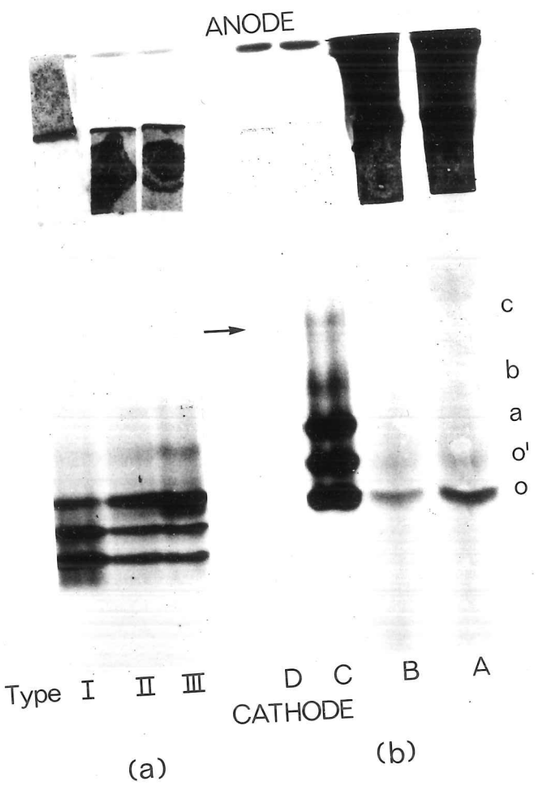


Fig 5.2
a) Examples of the extent of variation in relative intensities of the three main radioactive components of the isoelectric focusing pattern; b) Intracellular and extracellular radioactive components analysed by isoelectric focusing. (A) Clone giving type III pattern (D1). (B) Clone giving type I pattern (D2). (C) Extracellular control of 24 hr. labelling. (D) As (C) but with non-producer clone (NP2). The arrow marks the position of the strongest IgG band in a sample of P3K serum run under the same conditions.

into the cloned cells either by transferring the visible colonies to liquid medium and incorporating the cells in suspension, or by overlaying the soft agar with 1.4 mls (for a 60 mm Petri dish) of medium C (minus lysine) containing 10 µCi ^{14}C-Lys and 0.36% agar. The Petri dish was then incubated for a further 24 hr. Each colony in a plug of surrounding agar was transferred with a Pasteur pipette onto a polyacrylamide slab for isoelectric focusing.

5.5 Variation in Band Pattern

The secreted Ig of a number of clones derived from P3K was examined by cloning in agar and incorporating labelled amino acid in suspension. The pattern of radioactive Ig produced by each clone consisted of the same three major bands and so these cannot be the result of clonal heterogeneity. However the relative intensities of the three bands varied from clone to clone. Fig.5.2a shows two extreme pattern types and a typical intermediate pattern. The pattern type was a stable characteristic of each clone. Two clones exhibiting extreme pattern types (D1 and D2) were maintained in continuous culture for six months and on periodic testing the patterns remained relatively constant. When these clones were sub-cloned, however, the band pattern did not "breed true" i.e. on recloning a clone exhibiting any given pattern type, the daughter clones exhibited the full range of patterns.

The band pattern was shown to be unrelated to the isoelectric point of the newly synthesised Ig by labelling the cells for 15 min. with ^{14}C-Lys (see methods). Clones D1 and D2 produced an identical single major band as shown in Fig.5.2b. The conversion of this single band

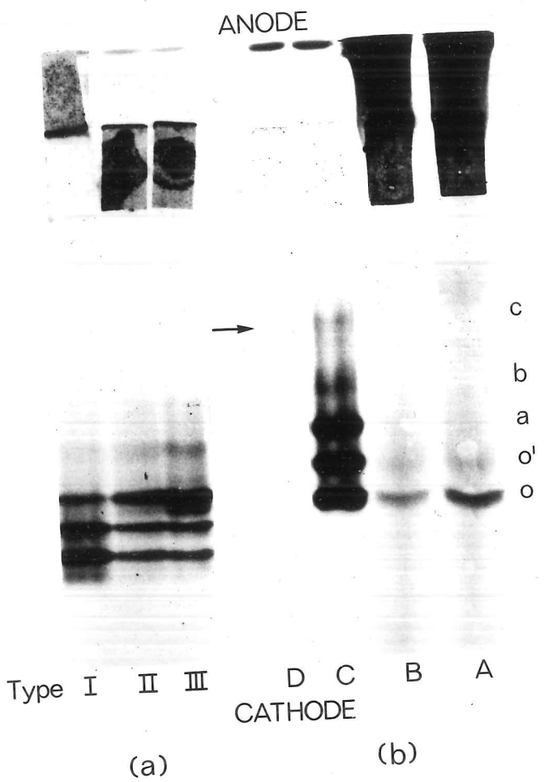


Fig 5.2
a) Examples of the extent of variation in relative intensities of the three main radioactive components of the isoelectric focusing pattern; b) Intracellular and extracellular radioactive components analysed by isoelectric focusing. (A) Clone giving type III pattern (D1). (B) Clone giving type I pattern (D2). (C) Extracellular control of 24 hr. labelling. (D) As (C) but with non-producer clone (NP2). The arrow marks the position of the strongest IgG band in a sample of P3K serum run under the same conditions.

into the cloned cells either by transferring the visible colonies to liquid medium and incorporating the cells in suspension, or by overlaying the soft agar with 1.4 mls (for a 60 mm Petri dish) of medium C (minus lysine) containing 10 µCi ^{14}C-Lys and 0.36% agar. The Petri dish was then incubated for a further 24 hr. Each colony in a plug of surrounding agar was transferred with a Pasteur pipette onto a polyacrylamide slab for isoelectric focusing.

5.5 Variation in Band Pattern

The secreted Ig of a number of clones derived from P3K was examined by cloning in agar and incorporating labelled amino acid in suspension. The pattern of radioactive Ig produced by each clone consisted of the same three major bands and so these cannot be the result of clonal heterogeneity. However the relative intensities of the three bands varied from clone to clone. Fig.5.2a shows two extreme pattern types and a typical intermediate pattern. The pattern type was a stable characteristic of each clone. Two clones exhibiting extreme pattern types (D1 and D2) were maintained in continuous culture for six months and on periodic testing the patterns remained relatively constant. When these clones were sub-cloned, however, the band pattern did not "breed true" i.e. on recloning a clone exhibiting any given pattern type, the daughter clones exhibited the full range of patterns.

The band pattern was shown to be unrelated to the isoelectric point of the newly synthesised Ig by labelling the cells for 15 min. with ^{14}C-Lys (see methods). Clones D1 and D2 produced an identical single major band as shown in Fig.5.2b. The conversion of this single band

to three takes place intracellularly within 1 hr. of synthesis. The nature of this conversion is still unknown but it is unlikely to be due to deamidation (Awdeh et al, 1970). There was no correlation between the pH (as judged by the colour of the medium) at the end of the incorporation and the type of pattern.

Thus the variation in band pattern appeared neither to be due to an artefact of handling nor is it under simple genetic control, since subclones showed such variation from the pattern of the parent clone. In view of these complexities the problem was not investigated further.

5.6 Non-production of Ig

A less frequent variation was found to be the total absence of Ig bands. One out of 20 subclones of a D1 culture that was 90 generations old was such a "non-producer". Similar clones were also isolated from agar plates by overlaying with agar containing sheep anti-P3K IgG antiserum and selecting colonies around which no precipitate developed (Coffino et al, 1972). Cellulose acetate electrophoresis and SDS gel electrophoresis of the ^{14}C-Lys labelled extracellular medium also failed to reveal Ig or Ig chains. Analysis of intracellular material from two non-producers (NP1, NP2) by SDS gel electrophoresis gave no evidence of Ig chains but the possibility that non-secreted Ig components are rapidly degraded within the cell was not excluded.

Table 5.1 Calculation of the rate of mutation to non-production

Clone	Number detected (a)	Number screened (b)	Average age (Generations) N
NP1) NP2)	2	351	150
NP4	1	20	90
NP5	1	89	150
NP7	1	122	300

Mutation frequency (f) = $\frac{M}{N}$, where M = $\frac{a}{b}$ = proportion of variants found in the population.

$$= \frac{5}{(351\times150)+(20\times90)+(89\times150)+(122\times300)} = \underline{4.8\times10^{-5}/\text{cell/generation}}$$

5.7 Frequency of mutation to non-producers

Five non-producers were detected in the early studies. Many more non-producers were observed in the routine screening (see below) but these were not re-examined and confirmed as non-producers and are not included in the calculation (Table 5.1) which gives the rate of mutation to non-producers as about 5×10^{-5}. The calculation takes into account the steady accummulation of mutant clones which takes place in a population, assuming that the mutations are selectively neutral. This assumption was not verified independently and so any mutation rates derived from it are minimum values. However the isolation of non-producer mutant clones from the population suggests that the loss of any function that the myeloma proteins might possess does not drastically increase the mean generation time of the cells. Non-production may arise from mutations in control genes, but most somatic theories of diversity require that variation in the structural gene product must also arise spontaneously without affecting the division rate of the cells. Are such events also detectable using the myeloma cell population as a model and, if so, what is their nature? To answer this a more extensive search for variant cells was performed.

5.8 Electrophoretic Variation

The method of incorporating radioactivity by overlaying colonies with agar containing ^{14}C-Lys was used to screen for less frequent forms of variation and in particular for clones secreting Ig of altered IEP. The reliability of the procedure in detecting such clones was tested by mixing P3K cells with those of a different cell line (P1). After cloning and incorporation of ^{14}C-Lys clones secreting P1 IgG could be clearly distinguished from those secreting P3K IgG,

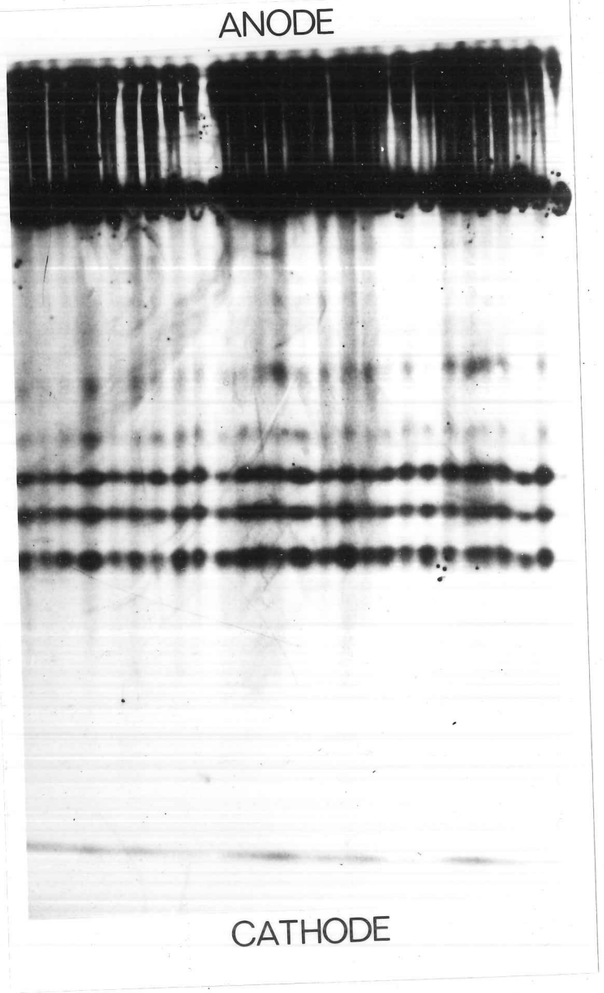


Fig. 5.3
Radioautographic screening of 27 clones from P3K spinner culture using original (destructive) method.

by virtue of the difference in isoelectric focusing pattern between the two proteins.

Screening of 1250 clones from the P3K spinner was then carried out. The culture was about 150 generations old when the experiment was started and about 250 generations old at the end. Of the colonies examined a number gave weak or negative results. Presumably the clones were small, dead or non-producers. These represented between 30% (in early experiments) and 4% (in later ones) of all those tested. Of the remaining 1102 whose band pattern could be positively identified, all but one produced the three major bands characteristic of P3K IgG (Fig.5. 3). It was not possible to study the clone that produced the variant pattern any further since the cells of that clone were destroyed during the analysis of the secreted protein. However the exciting observation of a variant pattern merited the extension of the method to allow the maintenance of replicate cultures whilst analysing the secreted protein.

5.9 Improved Screening Method

The most important modification to the screening was the introduction of a separate dish for the incorporation of labelled amino acids. Colonies were grown in soft agar as before, but transferred in their surrounding plugs of agar onto strips of dialysis tubing. ^{14}C-Lys, solidified in a dish of agar beneath the dialysis tubing, can diffuse through the membrane into the living colonies, whereas secreted protein remains trapped in the agar plug. After a 24 hr. incubation the protein is analysed by inverting each strip onto the origin of a

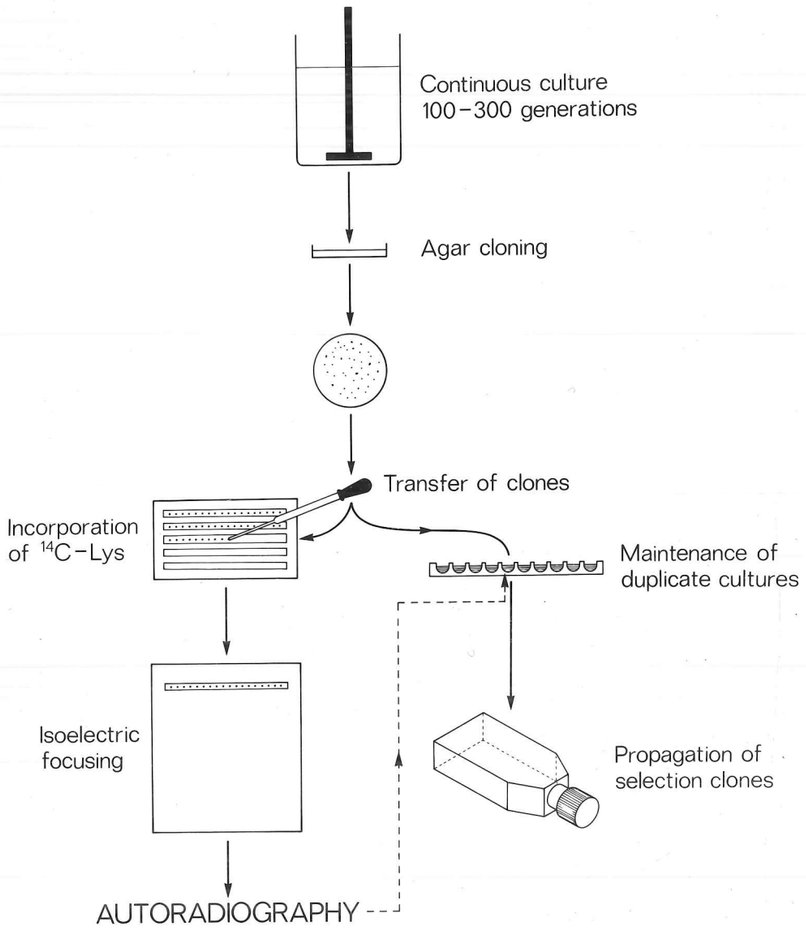


Fig. 5.4
Diagram of procedure used to isolate mutant clones of P3K.

polyacrylamide slab prepared for isoelectric focusing. Replicate cultures of each colony were established by inoculating micro-cultures with a few cells retained in the Pasteur pipette used for the transfer of the agar plug to the dialysis tubing. Each well in a microculture plate thus corresponds to an agar plug in the incorporation dish. The microcultures are maintained until the result of the isoelectric focusing is known and any clones which require further study are propagated from the microculture. The method is summarised in Fig. 5.4 .

A number of other modifications were incorporated into the improved procedure: 1) 6M urea was used in place of distilled water (to give a final concentration of 4M) in the composition of the gel mixture. This resulted in less radioactivity associated with insoluble material at the origin of the gel and a higher yield of radioactivity in the Ig bands. The addition of 100 µg of unlabelled P3K IgG to a 20 λ sample of extracellular medium was observed to have a similar effect, but urea was preferred for routine use. 2) In order to reduce the time required for screening large numbers of colonies, an apparatus for preparing and running several isoelectric focusing gels in parallel was designed Figs. 2.1-2. 12 gels of 20 samples each were routinely processed together. 3) Isoelectric focusing at 4^{o} instead of at room temperature resulted in sharper bands. 4) Strips (18x1 cm) of Whatman No.1 paper presoaked in the appropriate electrode buffer were used as wicks to ensure uniform electrical contact between the inverted gel and the carbon electrodes.

Using the improved procedure 6000 clones from a spinner of P3K cells were screened for variants. At the start of the experiment

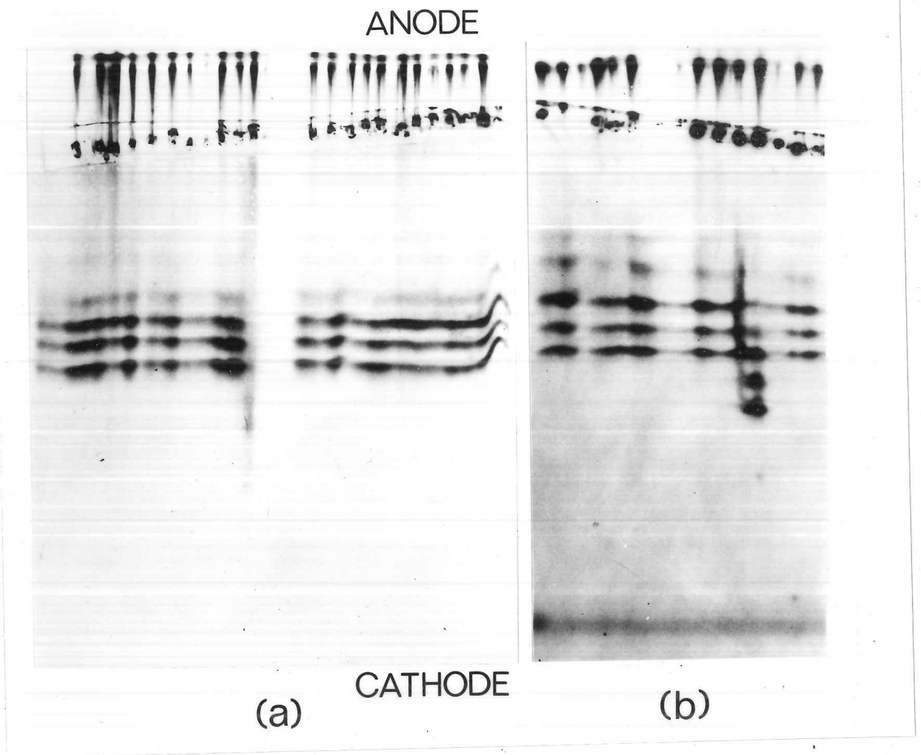


Fig. 5.5
Radioautographic screening showing detection of a) IF-1; b) a mutant clone that was not recovered.

the spinner had undergone about 100 generations since cloning, and at the end this had increased to about 400. The age of the culture could only be calculated approximately since the mean generation time was not constant throughout the life of the culture, but varied between about 18 and 36 hr. An average value of 24 hr. was estimated for calculations of mutation rates (see also Cowan & Milstein, 1974). 4875 colonies produced distinct band patterns. Whenever the radioautograph suggested the possibility of a protein of altered IEP, cells from the corresponding microculture were allowed to incorporate ^{14}C-Lys and the extracellular medium analysed by isoelectric focusing. In this way it was possible to eliminate as artefacts a number of irregular band patterns which arose.

In one radioautograph a new pattern was apparent (Fig. 5.5a) and this was reproduced on testing the microculture. In addition to the new bands there was a background which arose from cross-contamination of cells from neighbouring clones during the manipulations. Pure mutant clones were isolated by recloning (Fig. 6.2b). This clone, IF-1, produces none of the bands characteristic of the parental P3K IgG, suggesting that in these cells only a single allele is expressed. Thus the property of allelic exclusion (Pernis et al, 1965) is maintained in these cultured cells, even though they are known to be aneuploid (Horibata & Harris, 1970).

A second variant pattern was detected in further screening of the P3K spinner (Fig. 5.5b) but the replicate culture of this colony died and the clone could therefore not be propagated.

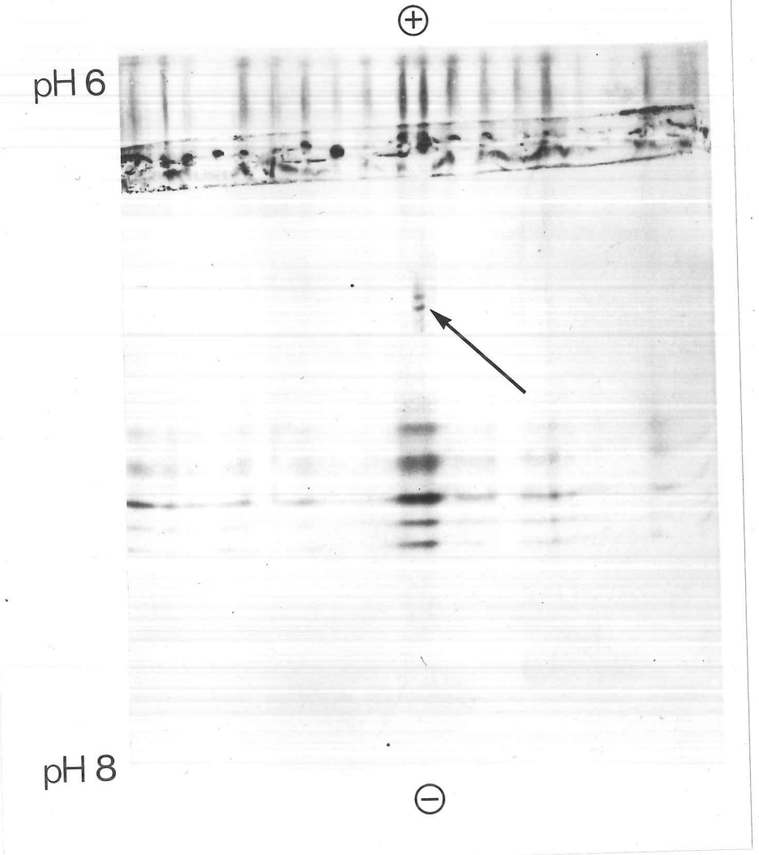


Fig. 5.6
Radioautographic screening. A suspected variant pattern is shown (arrow).

A third unusual pattern, very strongly suggestive of a mutant clone, was also seen and is shown in Fig. 5.6 . In this case the test incorporation of the microculture failed to reproduce the variant pattern, as did incorporations of sub-clones (which all produced the wild-type pattern). This result is normally indicative of an artefact in the analysis. However the striking resemblance of the variant pattern to that of a confirmed mutant protein (IF-2, see below) raises the possibility that the variant pattern did in fact arise from a clone of mutant cells but that 1) this clone was overgrown by cross-contaminating wild-type cells in the microculture, or 2) the sample of cells transferred from the agar plug to the microculture contained no cells representative of the mutant clone, but only cross-contaminating wild-type cells. Without viable cells producing the variant pattern, it was not possible to investigate these alternatives further.

A further variant clone (IF-2) was isolated from a different population of P3K cells which was about 450 generations old. This clone, after purification by sub-cloning, also expressed a variant band pattern which in this case, in contrast to IF-1, was of lower IEP than the wild type.

5.10 Frequency of Mutation

Of the 7117 clones which gave a positive result, two clones secreting a variant Ig were recovered and a further three probable variants were detected but not isolated. Two cultures were used,

both with an average age of about 300 generations since cloning. Using the equation M = f.N and making the assumptions discussed above, the frequency of "mutation to electrophoretic variation" is of the order of $\frac{2}{7117 \times 300}$ to $\frac{5}{7117 \times 300}$ i.e. $0.9\text{-}2.3 \times 10^{-6}$/cell/ generation.

The higher figure is in close agreement with the published rate (4×10^{-6}/cell/generation, Cotton et al, 1973) based on the detection of the first two variant band patterns.

5.11 Discussion

Limitations of Detection

In order to relate the frequency of electrophoretic (phenotypic) variation to the mutation rate in the structural gene coding for that protein it is necessary to assess the importance of two factors;

a) What proportion of mutations result in an altered phenotype?

b) What proportion of phenotypic variants are in fact due to a mutation in the structural gene coding for that protein?

5.11.1 Undetectable Mutations

A number of lines of evidence strongly suggest that any mutation giving rise to a single charge change in the Ig molecule will produce a detectable change in the isoelectric band pattern. It is generally assumed that mutations which result in the substitution of an amino acid residue for one of like charge (electrophoretically silent mutation) will not be detected by isoelectric focusing. Estimates of the proportion of point mutations which lead to

electrophoretic variants make certain assumptions about the *in vivo* assignment of degenerate codons and the uniformity of mutation frequency with respect to position within a codon. A value of 25-30% is often quoted (e.g. Shaw, 1965). A further class of mutants which may not be detected are those in which the mutation is deleterious and results in an increase in the mean generation time of the cells. The importance of this effect has not been measured. It seems likely however that in the case ofa myeloma protein, which has no function in the culture, such deleterious mutations will be less frequent than with an enzyme or other functional gene product with more precise sequence requirements.

5.11.2 Mutations in Structural Genes

In previous work on spontaneous mutation in animal cells it has not been possible to assess whether the phenomenon of electrophoretic variation reflects mutation in a single gene or the sum of mutations in a number of genes, including those specifying enzymes involved in post-translational modification of the polypeptide chain. For this purpose P3K cells are extremely suitable for mutational studies since it is possible to obtain large amounts of unlabelled wild-type and mutant protein from the serum of tumour bearing mice as well as radioactively labelled proteins of high specific activity from the tissue culture medium. In order to characterise the nature and location of the mutations resulting in IF-1 and IF-2, a chemical comparison of the variant and wild-type proteins was undertaken.

CHAPTER 6

Chemical Studies of IF-1 and IF-2 Immunoglobulins

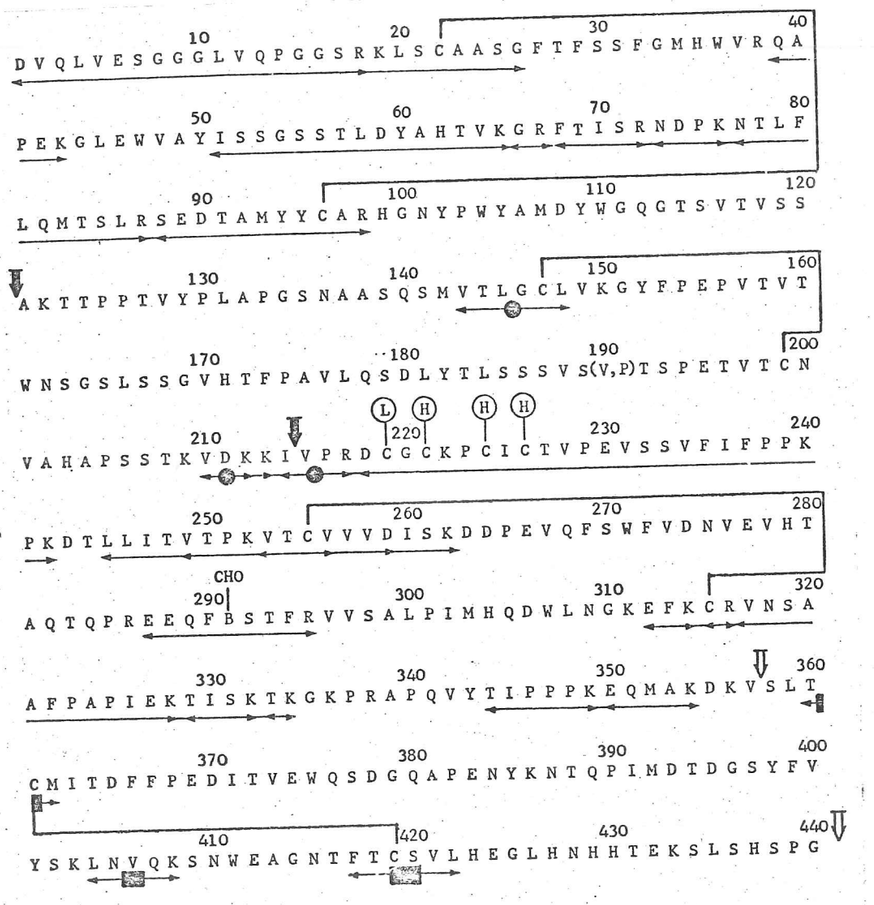


Fig. 6.1 Sequence of P3K H-chain (Milstein et al, 1974b) showing deletions in IF-1 and IF-2. Arrows under the sequence show the peptides that were identified by fingerprinting procedures, amino aicd analyses, N-terminal analyses and (in some cases) sequence determination. Tables 6.2 and 6.3 contain the key to the location of the peptides. Peptides marked ◄▬▬► and ◄●► were not detected in fingerprints of IF-1 and IF-2 respectively. Vertical arrows indicate the beginning and end of the deletions proposed for IF-1 (⇓) and IF-2 (⇓) on the basis of data described in this dissertation and on that obtained subsequently (Milstein et al, 1974b). CHO indicates the site of carbohydrate attachment. Lines between half-cystine residues indicate disulphide bridges.

CHAPTER 6

Chemical Studies of IF-1 and IF-2 Immunoglobulins

6.1 Introduction

In the previous chapter a technique was described which allowed the detection and isolation of two clones of P3K cells that secreted myeloma proteins of altered IEP. These variant cloned lines were grown up to mass culture and maintained continuously for several months. During this time no tendency to revert to production of the "wild-type" P3K IgG was observed. In order to determine whether this spontaneous and apparently stable variation was a) the consequence of a mutation in a structural gene and b) of a type that could be invoked in the generation of antibody diversity, a series of experiments was carried out to characterise chemically the differences between P3K and mutant IgG.

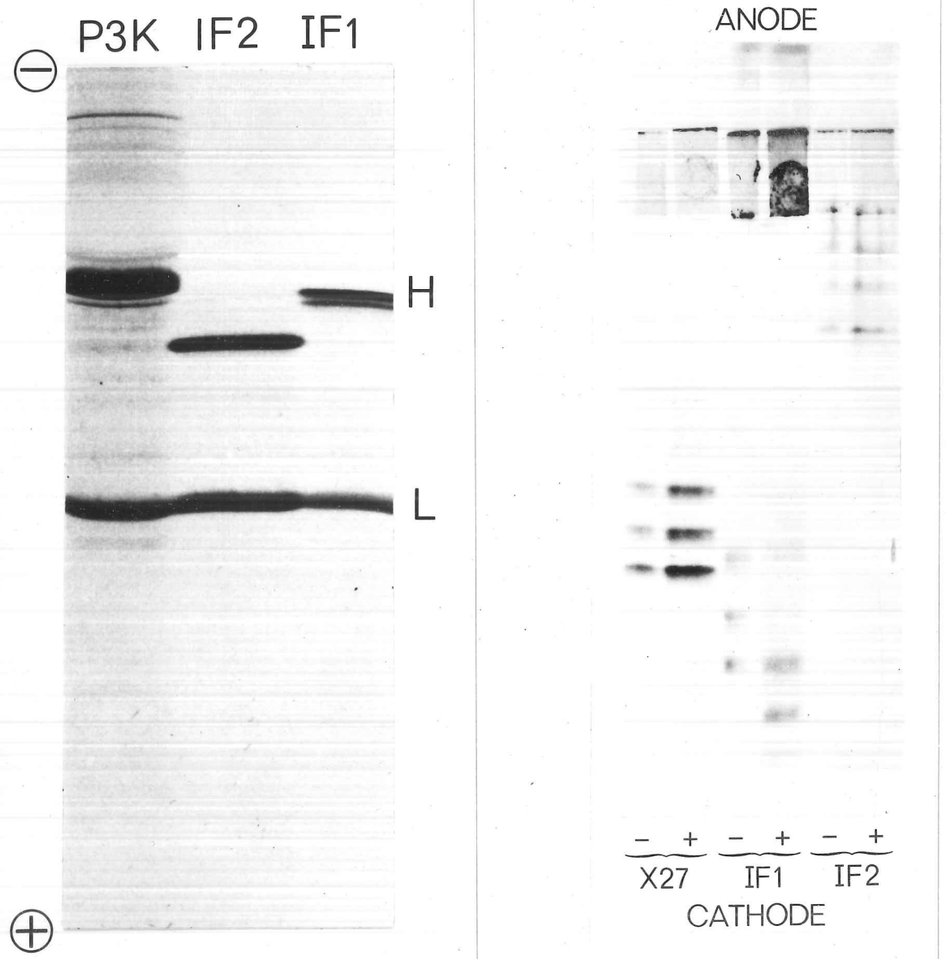
6.2 Use of Radioactive Protein

It is impractical to prepare sufficient unlabelled IgG for chemical studies from cells grown in tissue culture, since the secreted Ig in the growth medium is very dilute and heavily contaminated with horse Ig from the added serum. By contrast, radioactively labelled protein, specifically labelled with one or more amino acids can be prepared with much higher efficiency in tissue culture than in tumour-bearing mice or using tumour fragments (Askonas, 1961). Proteins may be labelled either externally, using a radioactive chemical modification reagent (e.g. ^{14}C-iodoacetic acid) or internally by the incorporation of radioactive amino acid or amino acid precursors, into the protein. This approach

is analagous to the very successful use of ^{32}P-labelled nucleoside triphosphates in sequence studies of nucleic acids (e.g. Jeppesen et al, 1970; Brownlee & Sanger, 1969). External labelling of proteins is in widespread use, but incorporation of labelled amino acids for sequence studies, although used in a bacterial system in the classic experiments on colinearity of the genetic code (Sarabhai et al, 1964), has been very little developed for sequence studies, and especially for proteins of eukaryotic cells. A major drawback for complete sequence studies is the difficulty of obtaining uniform labelling of all the amino acid residues, (see Sanger, 1958). However, specific labelling with one or more amino acids can provide very rapidly partial sequence information. This approach has been successfully exploited in the determination of the structure of the disulphide bridges of P3K IgG (Svasti & Milstein, 1972a).

The work described in this chapter uses radioactively labelled protein in two main ways. Firstly, as a means of detecting small amounts of protein (e.g. on gels) and secondly, in fingerprinting studies, as a means of examining for the presence of peptide derived from specific areas of the molecule. By using different labelled amino acids and/or changing the mode of digestion it was possible to examine various regions of the molecule for changes between wild-type and variant proteins. It was not possible to obtain detailed sequence information from the variant proteins using radioactive protein alone. However from the mobility of the modified peptides it was possible to locate quite precisely the position of the alterations.

For a complete description of the nature of the variations, sequence studies on unlabelled protein are required. Initially it



a b

Fig. 6.2
Comparison of wild-type (X27, P3K) and variant (IF-1, IF-2) ^{14}C-IgG by a) SDS-gel electrophoresis of H- and L-chains from purified IgG and b) by isoelectric focusing of crude extracellular medium. + indicates predigestion of the sample with neuraminidase (10 µg/ml, 4 hr., 37°).

was very difficult to obtain useful amounts of unlabelled variant myeloma protein. Subcutaneous injection of about 10^7 tissue culture cells into mice sometimes produced tumours and myeloma protein in the serum. Often these tumours regressed or were rejected on transfer to healthy animals. As soon as sufficient unlabelled IF-1 IgG was available, a diegest was performed in which the added radioactive material served now as a marker for location of the unlabelled, variant peptides in an efficient and non-destructive manner. It was thus possible to elute and sequence a peptide unique to IF-1.

The problems of production of unlabelled variant protein have now been overcome. After serial passage for several months of the few tumours that were not rejected at any one transfer and by irradiating the recipient mice before transfer of the tumours to suppress their immune response, the mutant tumour lines appear to have adapted to growth in mice and may now be passaged with the usual efficiency in healthy unirradiated Balb/c mice. Large amounts of unlabelled protein are now available for further characterisation of the variant proteins.

6.3 IF-1

6.3.1 SDS-Gel electrophoresis

Reducing SDS-gel electrophoresis showed that IF-1 secretes both H- and L-chains and thus can not be the consequence of a mutation resulting in the loss of synthesis of one of the chains (Fig.6.2a). The L-chain of IF-1 had the same mobility as that of P3K, but the H-chain appears to be smaller than its wild-type counterpart. The identity of the L-chains from P3K and IF-1 was also indicated by

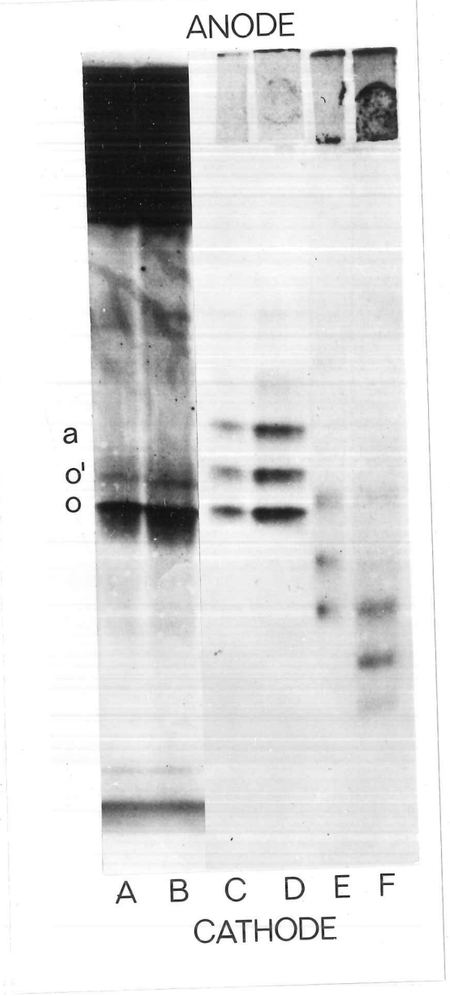


Fig. 6.3
Effect of neuraminidase digestion on P3K and IF-1 IgG. A) P3K intracellular material labelled in 15 min.; B) As A) but treated with neuraminidase (10 µg/ml for 4 hr., 37°); C) P3K secreted IgG from 24 hr. incubation with ^{14}C-Lys; D) As C) but neuraminidase; E) IF-1 secreted IgG from 24 hr. incubation; F) As E) but treated with neuraminidase.

fingerprinting the soluble tryptic peptides (see below) and by isoelectric focusing of separated H- and L-chains. The difference in mobility on SDS-gel electrophoresis between IF-1 and P3K H-chains is equivalent to a deletion of about 10 amino acid residues from P3K H-chain. However this interpretation is complicated by the presence of carbohydrate on the H-chain of IgG. Since the mobilities of glycoproteins on SDS-gels are sometimes anomalous (Bretscher, 1971) a number of experiments were performed to determine whether the difference is due to alterations to the carbohydrate moiety of the H-chain or to the polypeptide chain itself.

6.3.2 Carbohydrate Analysis

The higher IEP of IF-1 H-chain entails an overall loss of negative charge or gain in positive charge. One possible explanation for such a change would be a mutation resulting in the loss of sialic acid which is the terminal monosaccharide in many glycoproteins (Gottschalk, 1966). In order to investigate this possibility samples of radioactively labelled IgG were treated with neuraminidase, which specifically removes sialic acid from glycoproteins (Cassidy et al, 1965) When 50 µl extracellular medium containing ^{14}C-Lys labelled IgG was incubated for 4 hr. at 37° with neuraminidase (enzyme concentration, 10 µg/ml) and subjected to isoelectric focusing the result was as shown in Fig. 6.3 . The labelled material synthesised by P3K cells was unaffected by the treatment whereas each band of the secreted IF-1 IgG was converted to a more basic form, presumably due to the removal of sialic acid from the H-chain. Intracellular IgG from both P3K and IF-1 (not shown) labelled for 15 min. with ^{14}C-Lys was also unaffected. Thus the difference between IF-1 and P3K IgG can not be due to the loss

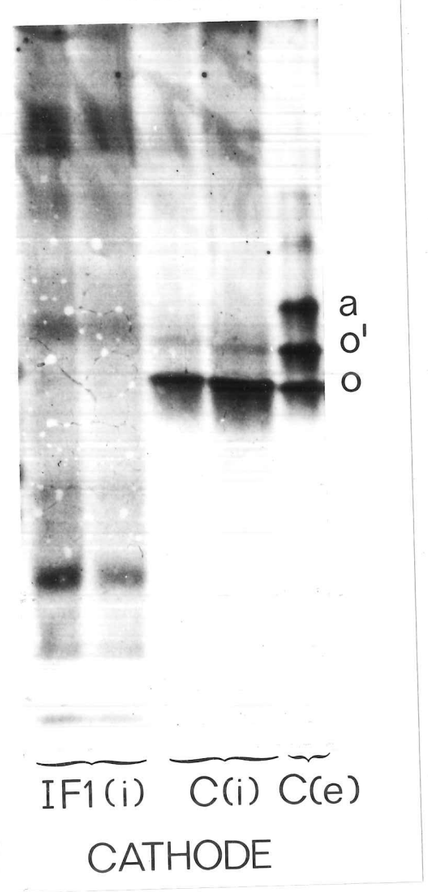


Fig. 6.4 Intracellular material of IF-1 and P3K cells was prepared as in section 2.6 and analysed by isoelectric focusing and radioautography. IF1(i) is intracellular material from IF-1 cells; C(i) from P3K cells. C(e) is secreted P3K IgG prepared according to section 5.2.1.

of sialic acid; indeed IF-1 appears to contain sialic acid whereas P3K does not and this interpretation was confirmed by radiochemical analysis of the carbohydrate contained in P3K and IF-1 IgG (Cowan et al, 1973). Each of the three major radioactive bands of IF-1 IgG was found to contain approximately equal amounts of sialic acid whereas none of those from P3K contained any. Three conclusions were drawn from these results:

1) Sialic acid cannot be involved in the recognition of P3K IgG as a protein destined to be secreted.

2) The difference between the three major bands cannot be due to different sialic acid content.

3) The mutation responsible for the altered properties of IF-1 IgG is not one which results in a simple alteration in carbohydrate addition. Sialic acid is gained not lost and yet the IEP of IF-1 is higher, not lower as would be expected if this were the only alteration. Thus the modification in carbohydrate structure must be in addition to, or the consequence of, an alteration in the polypeptide of the H-chain.

6.3.3 Intracellular IgG

The material from the extracellular medium as examined above contains secreted molecules as old as 24 hr. During this time it is likely that the protein is subject to other post-translational modification (e.g. deamidation of Asn and Gln residues) in addition to those affecting the carbohydrate moiety. In order to investigate the possibility that differences in such modifications could be responsible for the variant properties of IF-1 IgG, the newly synthesised proteins from mutant and wild-type cells were compared.

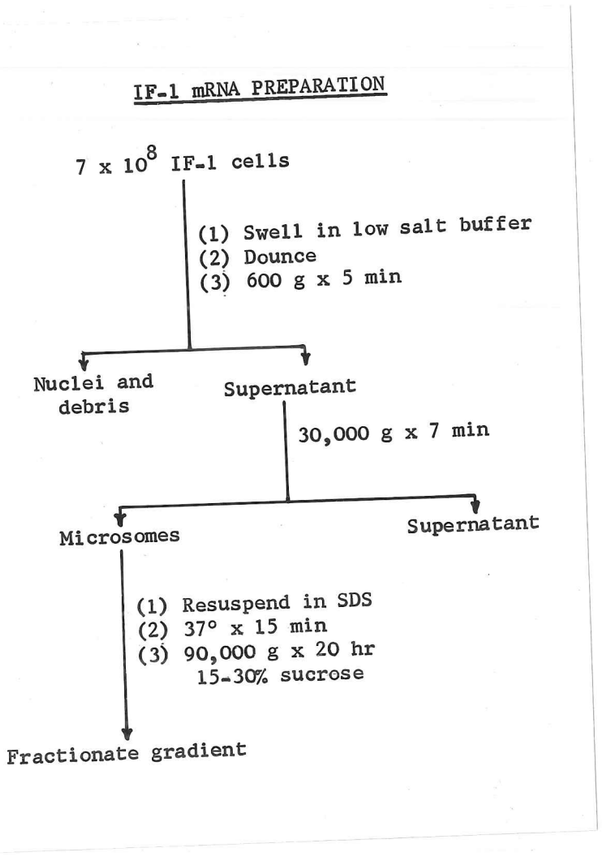


Fig. 6.5

Summary of protocol for preparation of Ig mRNA from IF-1 cells. See section 2.8 for details.

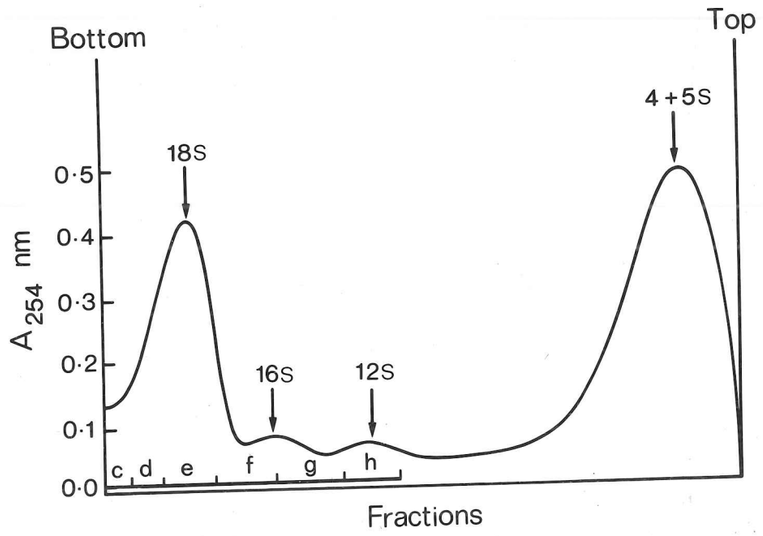


Fig. 6.6 Absorbance profile of microsomal RNA fractionated on a sucrose density gradient as described in section 2.8. The major peaks contain tRNA and 18s rRNA. The minor peaks are mitochondrial contaminants. L-chain mRNA is found in the 12-14s region, H-chain mRNA in the 18s peak (Brownlee et al, 1973; Cowan & Milstein, 1973). c-f refer to the fractions used to prepare IF-1 mRNA and correspond to the letters in Fig. 6.7.

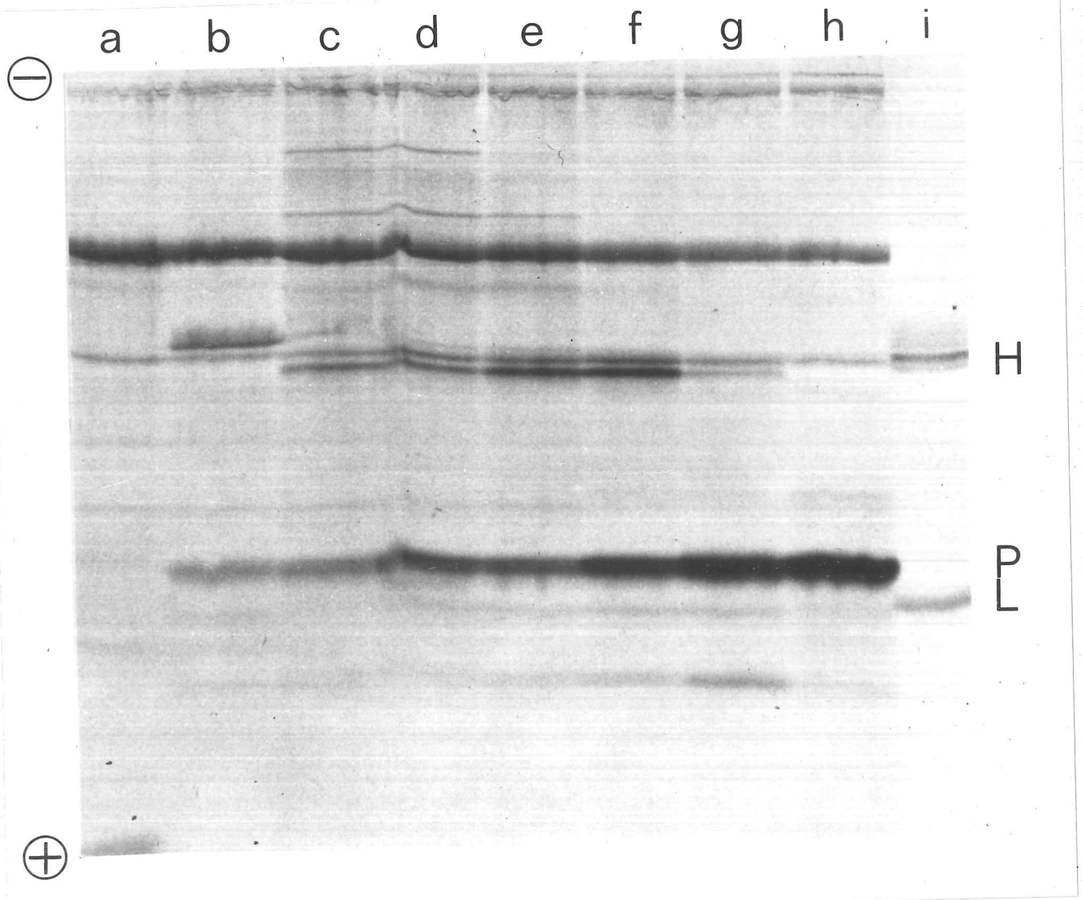


Fig. 6.7 Labelled products obtained by SDS-gel electrophoresis of rabbit reticulocyte lysates incubated with ^{35}S-Met. RNA was added to the incubations as follows: a) no RNA added; b) P3K H- and L-chain mRNAs; c)-h) IF-1 microsomal RNA of decreasing sedimentation coefficient from about 20S to 12S isolated from consecutive fractions of an SDS-sucrose density gradient; i) a mixture of ^{14}C-labelled IgG secreted by P3K and IF-1. Band P is the L-chain precursor (Milstein et al, 1972).

Using a short labelling time of 15 min. and examining the intracellular material as described in the previous chapter, it was possible to show that in a population of IgG molecules of which the oldest was not more than 15 min., there was a clear difference between the labelled material from IF-1 and P3K cells (Fig. 6.4). Thus any effect due to the addition of sialic acid, secretion, or other post-translational modification that takes place more than a few minutes after completion of the synthesis of the polypeptide chain, is eliminated as a possible cause of difference between P3K and IF-1 IgG.

6.3.4 mRNA

The above results implied that the mRNA coding for IF-1 and P3K H-chains are different. Purification procedures for Ig mRNAs are available (Brownlee et al, 1973; Mach et al, 1973 ; Swan et al, 1972) and also methods of assaying Ig mRNA by their ability to direct the translation of Ig chains in heterologous cells (Stevens & Williamson, 1972 , Smith et al, 1973) and cell-free systems (Stavnezer & Huang, 1971; Brownlee et al, 1972). To test the deduction that IF-1 and P3K mRNAs differ, they were compared in a cell-free assay.

Partially purified Ig mRNA was prepared from 8×10^8 IF-1 cells by a simplified version of the method of Brownlee et al (Brownlee et al, 1973). The protocol, which is described in detail in chapter 2 , is summarised in Fig.6.5 and involves the fractionation of microcosmal RNA on a sucrose density gradient (Fig.6.6). The mRNA contained in this preparation was compared with mRNA coding for H- and L-chains of P3K IgG (a gift from Dr T. M. Harrison) by observing their effect on the stimulation of protein synthesis in a cell-free system derived from rabbit reticulocytes.

Table 6.1

Calculation of IF-1 H-chain N-terminal Sequence

Cycle	1	2	3	4	5	6	7	8
IF-1 H+L	Asn Asp	Val Ile	Val Gln	Leu[b] Gly	Val Pro/Thr[c]	Gln Glu	Ser	Gly Pro
P3K L (Svasti & Milstein, 1972b)	Asn	Ile	Val	Met	Thr	Gln	Ser	Pro
Deduced IF-1 H	Asx[a,d]	Val	Gln	Leu	Val	Glx[a,d]	Ser[d]	Gly
P3K H	Asp	Val	Gln	Leu	Val	Glu	Ser	Gly

a In mixture analysis it is impossible to assign amides to both sequences with confidence.

b Met is destroyed on performic acid treatment. Gly is presumably a degradation product.

c Pro and Thr were not resolved from each other.

d The possibility of a basic residue or one labile to performic acid was not eliminated.

As shown in Fig. 6.7, RNA fractions from the 12-20S region of a sucrose gradient were observed to stimulate the incorporation of ^{35}S-Met into H- and L-chains. As expected the major product of synthesis directed by IF-1 mRNA of lower sedimentation coefficient co-electrophoresed with the translation product of P3K L-chain mRNA (Band P). However the protein synthesised on addition of faster sedimenting IF-1 mRNA (from about 18S) was markedly different from P3K H-chain (Fig.6.7).

These results indicate that the mRNAs coding for IF-1 and P3K H-chains are different and that the mutation involves an Ig structural gene. As a consequence of the mutation IF-1 H-chain appears to be smaller than P3K H-chain. The site of the apparent deletion in the IF-1 H-chain was investigated by N- and C-terminal analysis of unlabelled protein purified from the serum of mice bearing IF-1 tumours and by comparison of peptide fingerprints of radioactively labelled protein.

6.3.5 Automatic N-terminal analysis

Performic acid oxidised IF-1 IgG was subjected to 8 cycles of automatic Edman degradation. Amino acid derivatives from both H- and L-chain N-terminal sequences were produced, but by subtracting the known L-chain sequence, it was possible to deduce a partial H-chain sequence (Table 6.1). This sequence was identical to the N-terminal sequence of P3K H-chain, determined in a separate experiment. The possibility of an N-terminal deletion was therefore eliminated.

6.3.6. C-terminal analysis

The C-terminal of IF-1 H-chain was determined by carboxypeptidase

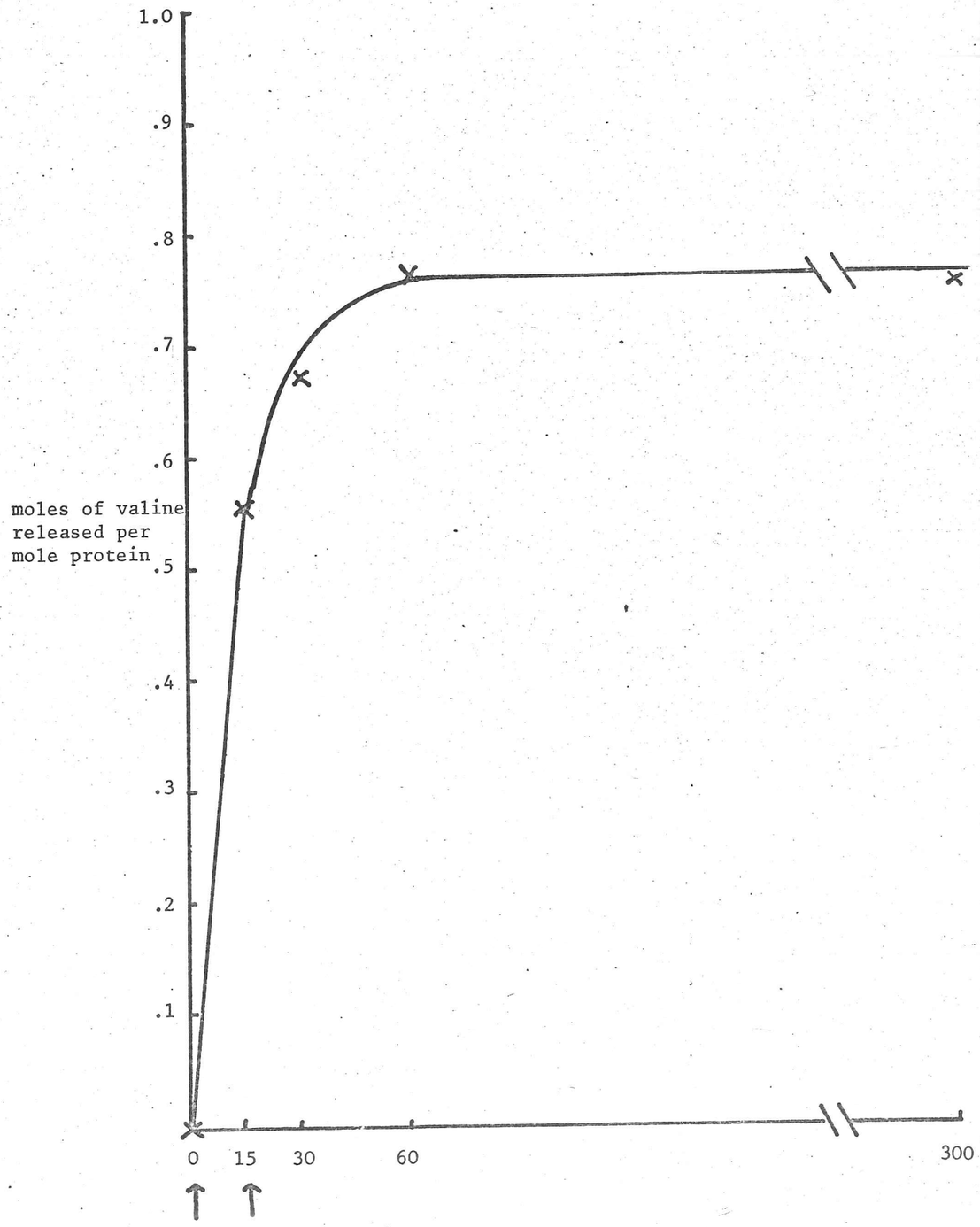


Fig. 6.8

Time-course of release of free amino acids by digestion of IF-1 H-chain with carboxypeptidase A (added at T=0) and carboxypeptidase B (added at T=15 min.). Samples were taken at different times and the release of amino acids determined on an amino acid analyser.

digestion. Carboxypeptidase A released Val from IF-1 H-chain (Fig.6.8). No other amino acid was released on subsequent addition of carboxypeptidase B. No free amino acid was released from P3K H-chain by a mixture of carboxypeptidases A and B, as expected from the known C-terminal sequence -Pro-Gly (Fig. 6.1). These results therefore suggested that the C-terminal section of the wild-type H-chain is missing in IF-1. To assess how far this deletion extends into the H-chain, a comparison of the soluble tryptic peptides of P3K and IF-1 H-chains was undertaken.

6.3.7 Fingerprinting of Tryptic Peptides

Tryptic fingerprints of labelled IF-1 H-chain were performed in two ways. With added unlabelled P3K IgG it was possible to identify accurately those peptides missing or altered in the variant protein. After a second digest in which unlabelled IF-1 H-chain was added, a variant peptide was purified and analysed. This information could not be obtained from the first digest since peptides unique to the variant protein were present in only radiochemical amounts.

^{14}C-labelled P3K and IF-1 H-chains were separately fingerprinted by conventional methods (Fig.6.9). Unlabelled P3K H-chain was added to each sample before digestion to serve as an internal standard for comparison of the two fingerprints. Many of the resulting carrier peptides were eluted and analysed and, by comparison with the known sequences of P3K H-chain (Fig. 6.1) most spots could be assigned to their position in the H-chain.

1.5×10^6 cpm ^{14}C-Lys, ^{14}C-Arg labelled IF-1 H-chain was added to 15 mg carrier P3K H-chain oxidised with performic acid and digested

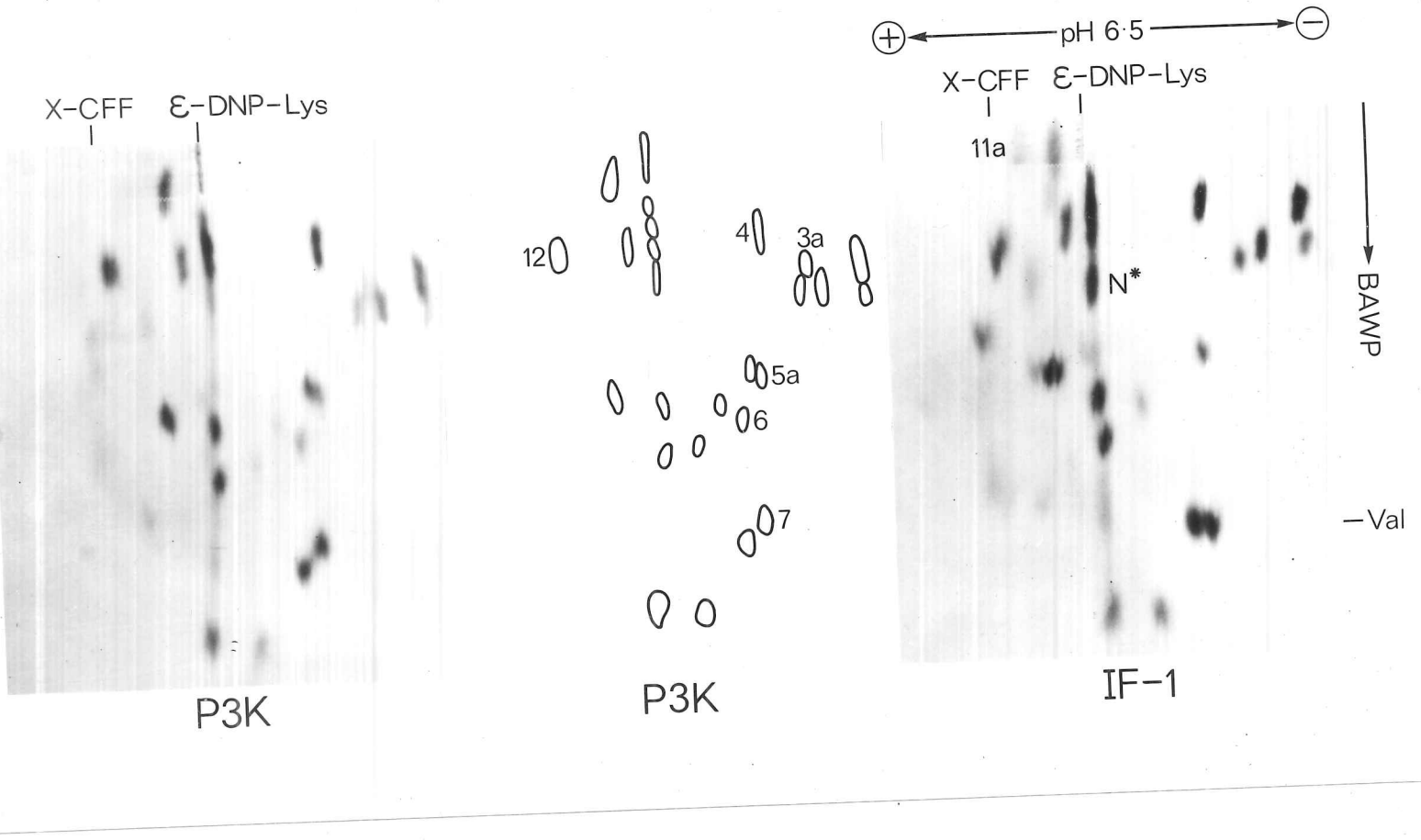


Fig. 6.9
Comparison of tryptic fingerprints (radioautographs) of ^{14}C-Lys, ^{14}C-Arg labelled P3K and IF-1 H-chains.
For compositions and location of peptides see Table 6.2 and Fig. 6.1.

Peptide	1a	2	3	4	5a	5b	6	6b	7	8	9	10	11	12α	12β	12b	15	16-17	18a	N1	N2	N3A	N3B	N4	N5	N6	N*
Purification	6.5,B	6.5,B 6.5	6.5,B	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 6.5	6.5,B 8.9	8.9 0.9	3.5 1.0	6.5,B 3.5	6.5,B 3.5	0.9	6.5 6.5	6.5 0.9	6.5 +	6.5,B 3.5	6.5,B 3.5	6.5,B 3.5	S.65 B,3.5	6.5,B 3.5	6.5,B 3.5	6.5,B 3.5	S,6.5 B,3.5
Cys (O_3H)											1.0					0.9								0.9			1.1
Met (O_2)				1.0	1.0			1.5			1.4	1.0		1.3		1.6	1.5	1.0	3.8	1.1			1.5		1.9		
Asp			1.0			0.8	1	1.5		1.0	2.1	1.0		1.3		1.0		1.0	1.7								
Thr						0.9		3.4		1.2	1.8	1.0		2.4		1.3	2.1	1.0	1.7	1.1						1.6	
Ser					1.0						1.6	3.1		1.7	2.0	2.2	3.5		4.8	1.2	0.9	1.7		2.0	1.0	1.2	
Glu									1.0					5.0	1.0		1.3			1.9		1.1					
Pro							3							1.9			4.9									1.2	
Gly		1.4						1.4							1.0	2.2				2.8		1.1		1.1			1.1
Ala								0.8						3.1			3.0	2.7	3.9	1.1			1.0				
Val				1.1	1.0			1.4	0.9					1.9				0.8		1.0							
Ile						1.0	1	0.9	0.7	1.0							1.8										
Leu					0.9			1.0			3.0					1.8				0.9	1.0						
Tyr								0.8		1.0	1.1	2.0		2.0													
Phe																											
Trp														3.0	0.7			0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Lys	1.0		1.0	1.8	1.0	1.0	1	1.0 1.2								1.1	1.0		0.6								
His									1.0	1.1	1.0	1.0	1.0														
Arg		1.0																									
m (pH6.5)	-0.8	-0.65	-0.55	-0.42	-0.4	-0.36	-0.33	-0.18	-0.42	-0.36	-0.2	0.1	0.08	0.34	0.34	0.4	0.5	0.24	0.35	0	0	0	0	0	0	0	0
Yield (%)	15	2	11	42	10	17	1	5	5	20	35	34	20	12	13	10.5	30	3.6		14	14	3		9.1	11.7	2.1	-
N-terminal		Gly	Thr	Val	Leu	Thr	Thr	Ile	Ile	Phe	Asx	Glx		Asx	-	Ser	Asx	Val	Val		Glx	Glx	Val	Glx	Asx	-	Asx
Location	213	66-67	333 -334	210--212	404-408	329-332	344-349	51-65	214-217	68-72	77-87	287-295	315-316	218-242	39-43	88-98	1-18	253 -262		317-328	312-314	39-43	210-212	350-354	73-76	39-43	355-357

Table 6.2 The amino acid composition (residues/mole peptide), yields (expressed as moles peptide recovered per mole protein used) and N-termini of tryptic peptides of P3K H-chain. The best method of purification is also shown. Paper electrophoresis is indicated by the pH used, paper chromatography by B, gel filtration (Sephadex G-50) by S. Electrophoretic mobilities (pH 6.5) are expressed relative to Asp. Location refers to the position of the peptide (Fig. 6.1). N* was isolated from digests of IF-1 only.

with trypsin. A parallel digestion of 1.5×10^6 cpm ^{14}C-Lys, ^{14}C-Arg labelled P3K H-chain mixed with 5mg P3K H-chain was performed under the same conditions and both digests fractionated on paper. The labelled IF-1 fingerprint was stained with 0.1% ninhydrin and some of the carrier (P3K) peptides thus located, eluted and analysed.

The labelled P3K fingerprint was stained with Cd-ninhydrin. All the spots visible after staining coincided with the radioactive spots seen in the radioautograph (except for one neutral, non-radioactive peptide, presumably the C-terminal peptide of P3K H-chain). This confirmed the identity of the soluble tryptic peptides of P3K H-chain isolated from mouse serum and from labelled P3K IgG produced in tissue culture. The IF-1 fingerprint contained two radioactive peptides (N*, 11a) absent in P3K H-chain (as judged by visual comparison of the two fingerprints and by the negative amino acid analysis of the eluted spots). Peptide 5a (residues 404-408), absent in IF-1 is derived from the C-terminal domain of P3K H-chain. Peptide N4 (350 - 354) is common to P3K and IF-1 and thus, assuming the mutation did not result in multiple alterations in protein structure, the sequence between residue 354 and the N-terminus is presumably identical in P3K and IF-1 H-chains. Certainly all the other major soluble tryptic peptides identified were common to both fingerprints. The sequence of N* was obtained from a tryptic digest of ^{14}C-Lys, ^{14}C-Arg labelled IF-1 H-chain (5×10^6 cpm) mixed with 30 mg IF-1 carrier H-chain. The digest was fractionated on Sephadex G-50 and the small peptides separated on paper as above. Peptide N* was obtained in good yield and its sequence determined as Asp-Lys-Val. Since the C-terminus of the peptide is not Lys or Arg (nor a good site for chymotryptic cleavage) this must be the C-terminal peptide of IF-1 H-chain. Since the region between residue 354 and the C-terminus of P3K produces no

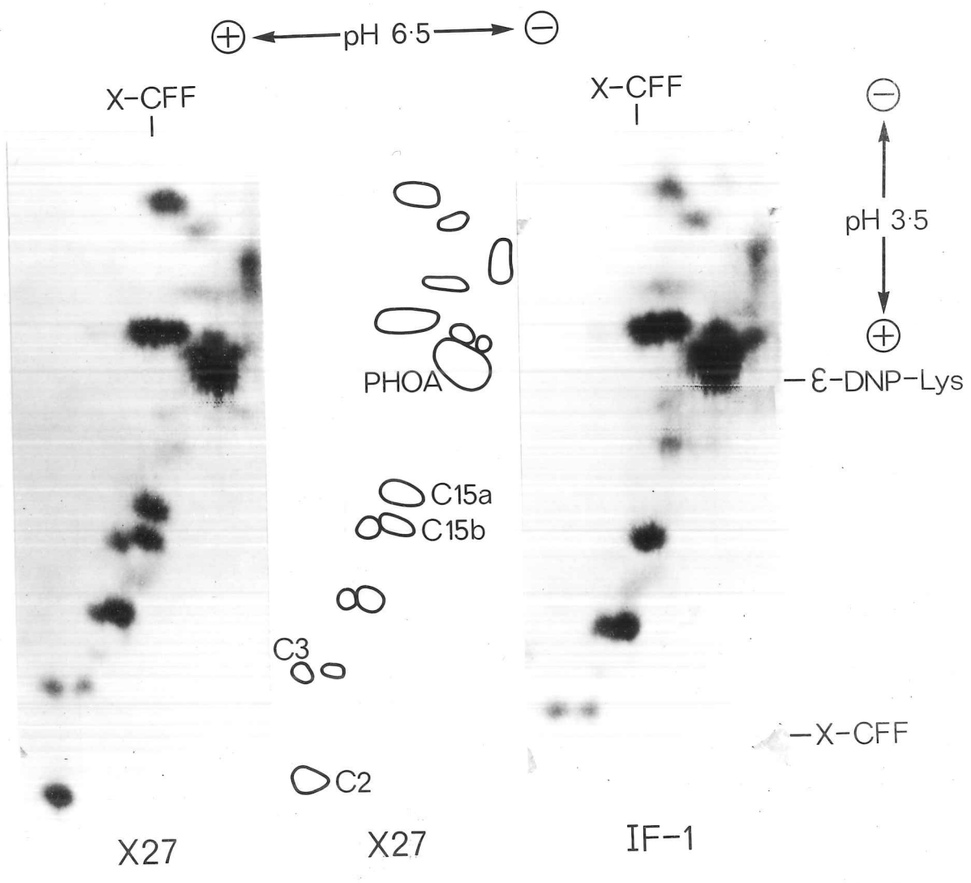


Fig. 6.10
Comparison of peptic fingerprints (radioautographs) of ^{35}S-Cys labelled P3K (x27 is a wild-type clone of P3K) and IF-1 IgG. Only the acidic peptides are shown here. For composition of the peptides see Table 6.3.

Table 6.3 The amino acid compositions (residues/mole peptide), and yields (expressed as moles peptide recovered/mole protein used) of peptic peptides of P3K IgG. The purification used is also shown: paper electrophoresis is indicated by the pH used, and paper chromatography by B. The deduced origin of the peptide is also shown by reference to the half-cystine it contained. I, II, etc. refer to H-chain i, ii, etc. L-chain, numbered from the N-terminus. (See also Fig.6.1)

Peptide	C2	C3	C4	C5	C6	C7	PHOA	C10b	C11	C13	C14	C15a	C15b
Purification	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B	6.5,B
	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5,2.1	3.5	3.5,2.1	3.5	3.5	3.5
Cys (O_3H)	1.0	1.0	0.9	1.0	0.9	1.0	4.5	0.9	1.0	0.6	0.8	1.0	1.0
Met (O_2)	1.0												
Asp						2.2	4.5		1.3		1.1		
Thr	1.0				0.9		3.4		1.7	3,0		1.0	0.9
Ser					1.0		3.3	3.4			1.0	0.9	
Glu					1.4	1.4	2.0	1.8					
Pro							5.7	1.0	1.9				
Gly		1.1					1.6	5.7					1.1
Ala					1.2		2.2	4.8			2.0		
Val			1.0	1.0			6.0	1.7	2.9	2.0	1.0	1.0	1.0
Ile							1.8			1.0			
Leu		1.0						1.8		1.2		1.0	2.1
Phe			1.1	0.9		1.0					0.9	0.9	
Lys					1.0		4.5	1.0	1.0	1.0	1.0		
His							1.2						
Arg						1.0	0.9	1.4			0.9		
Yield (%)	25	6	5	32	24	20	6	7	8	3	11	24	22
Cys	XI	III	iii	iii	i	v	IV-VIII	I	IX	IX	X	XII	III

small tryptic peptide other than 5a, the extent of the deletion was further investigated by comparing the peptides involved in disulphide bridges. Alignment of these peptides was by comparison with the known sequences around the disulphide bridges of P3K IgG (Svasti & Milstein, 1972 a).

6.3.8. Fingerprinting of peptic peptides

^{35}S-Cys labelled P3K IgG ($3x10^6$ cpm) and IF-1 IgG ($5x10^6$ cpm) were each digested with pepsin in the presence of 5mg carrier P3K IgG in 5% HCOOH. In Fig.6.10 the fingerprints are compared. Some of the peptides were further purified and analysed as indicated in Table 6.3. The major difference is the absence of peptides C15a and C2. Elution and analysis of these two peptides from a preparative digest allowed their identification as the half-cystine containing peptides involved in the C-terminal intrachain disulphide bridge. Thus, since residue 360 is absent in IF-1 though (from the tryptic fingerprints) 354 is present, the deletion must begin between 355 and 359. Other peptides were eluted from the preparative fingerprints and analysis of these showed that all the other P3K H-chain half-cystine containing peptides are present in IF-1.

6.3.9 Conclusion

The deletion of about 80 residues suggested by the above results is somewhat surprising in view of the mobility of IF-1 H-chain on SDS-gel electrophoresis, which indicated a M.W. difference of about 1,000 between P3K and IF-1. However this estimate is subject to the inherent uncertainties of glycoprotein M.W. determination by SDS-gel

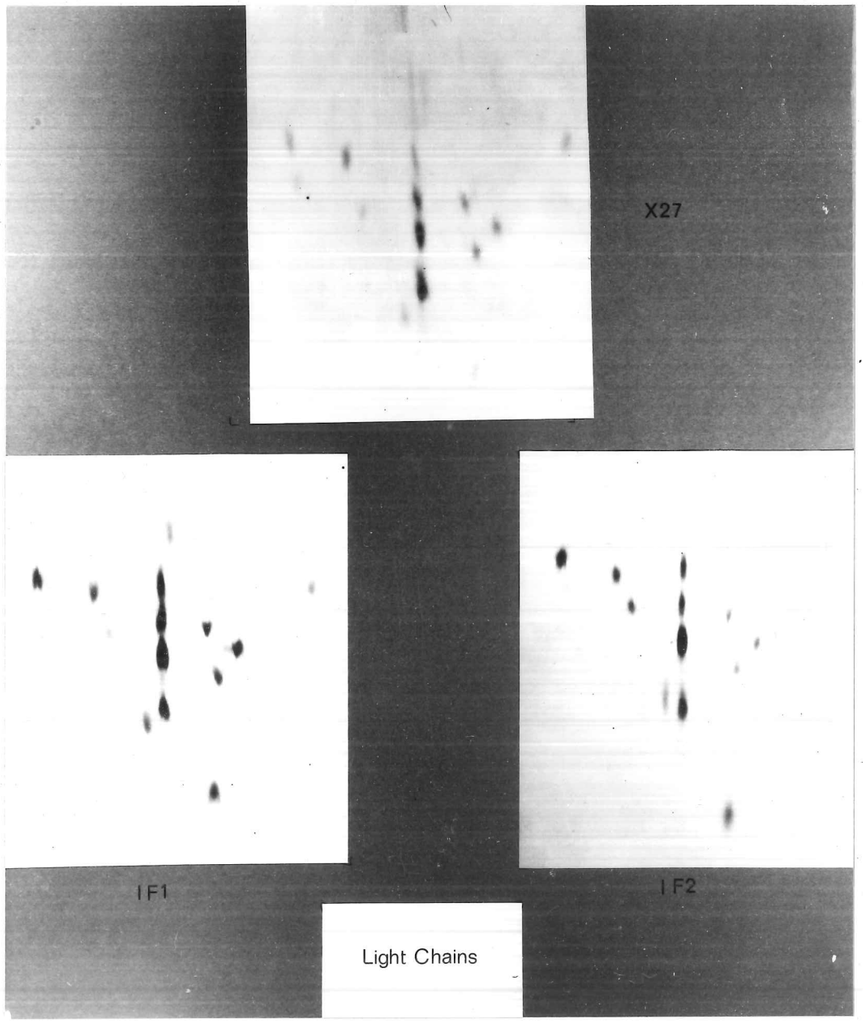


Fig. 6.11
A comparison of the tryptic fingerprints of wild-type (x27) and mutant (IF-1, IF-2) L-chains. For details of the fractionation see Fig. 6.9.

electrophoresis and especially since the carbohydrate moieties are known to be different. A more accurate comparison was obtained from the mobilities of the carbohydrate-free polypeptides synthesised in the rabbit reticulocyte lysate cell-free system (Fig.6.7). These suggested a M.W. difference of about 4,000 which is closer to the predicted difference. It is likely that the Val released by carboxy-peptidase A is the Val residue 357. The fact that -Asp-Lys-Val- is also the sequence of residues 355-357 in P3K H-chain (Fig. 6.1; Milstein et al, 1974b) indicates that this is the position of the C-terminus of IF-1 and that the mutation results in the deletion of all residues from position 358 to the C-terminus. The possibility that the missing residues are in part replaced by a new sequence which, coincidentally ends in (Lys or Arg)-Asp-Lys-Val is not, of course, excluded, but seems unlikely.

6.4 IF-2

A similar approach was adopted to determine the nature of the alteration in IF-2 IgG. SDS-gel electrophoresis (Fig.6.2a) suggested that IF-2 H-chain was missing about 80 amino acid residues and that the L-chain was again normal. Neuraminidase digestion and carbohydrate analysis (Cowan et al, 1973) did not reveal a difference between the carbohydrate of P3K and IF-2 and examination of intracellular IgG suggested that post-translational proteolysis is not responsible for the lower pI and greater mobility on SDS-gels of IF-2 H-chain.

6.4.1 Fingerprints of Tryptic Peptides

IF-2 L-chain fingerprint was identical to that of P3K, Fig. 6.11. The

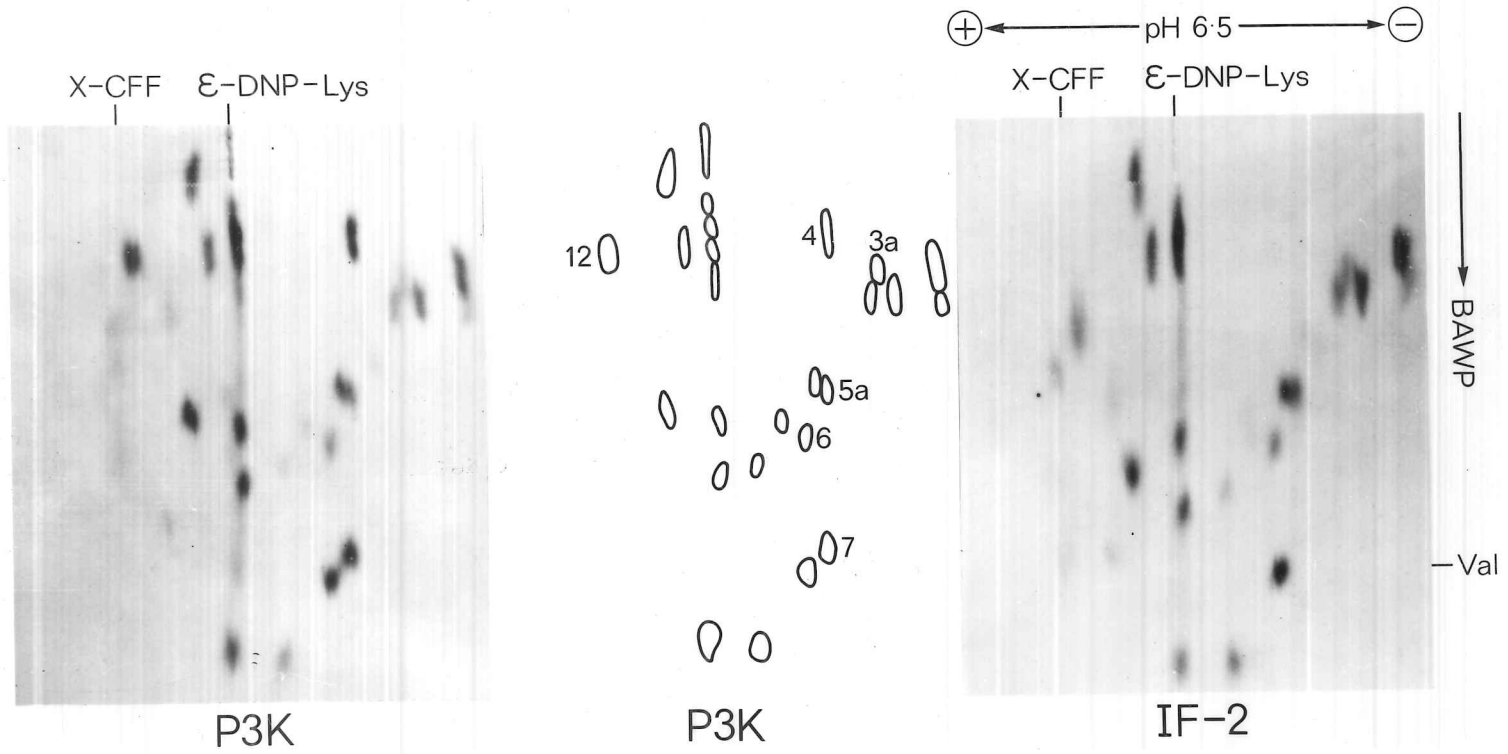


Fig. 6.12

A comparison of the tryptic fingerprints (radioautographs) of P3K and IF-2 H-chains. See Fig. 6.9 for complete key to peptides. For compositions and locations of peptides see Table 6.2 and Fig. 6.1.

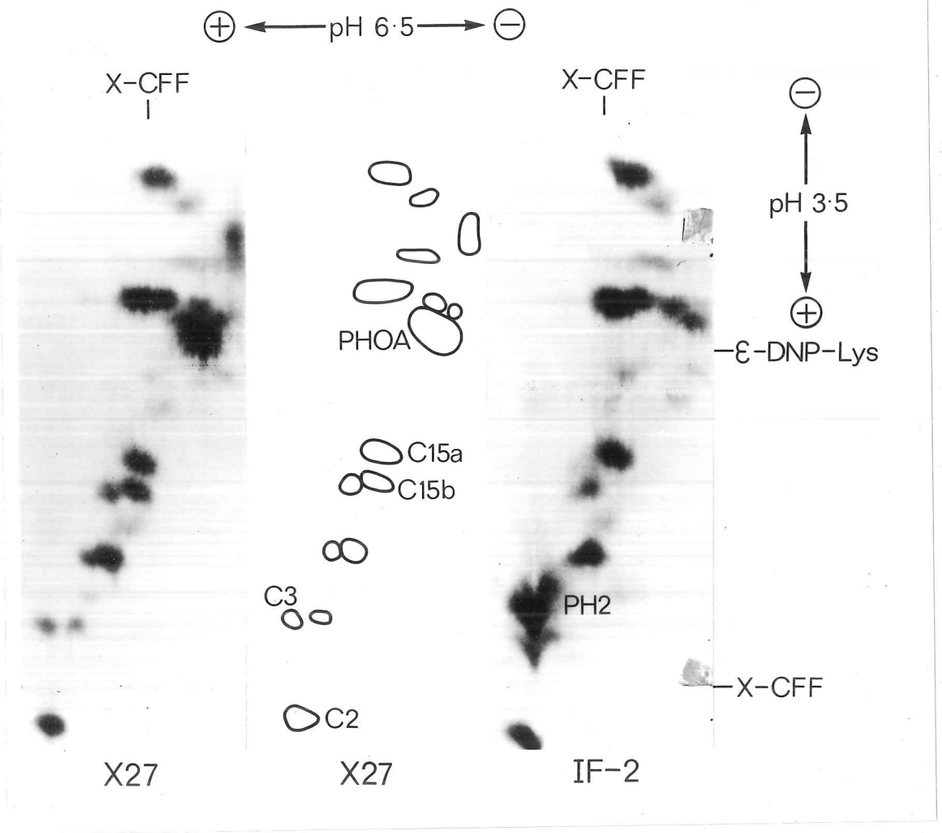


Fig. 6.13
Comparison of peptic fingerprints (radioautographs) of ^{35}S-Cys labelled P3K and IF-2 IgG. (X27 is a wild-type clone of P3K). For compositions of peptides see Table 6.3.

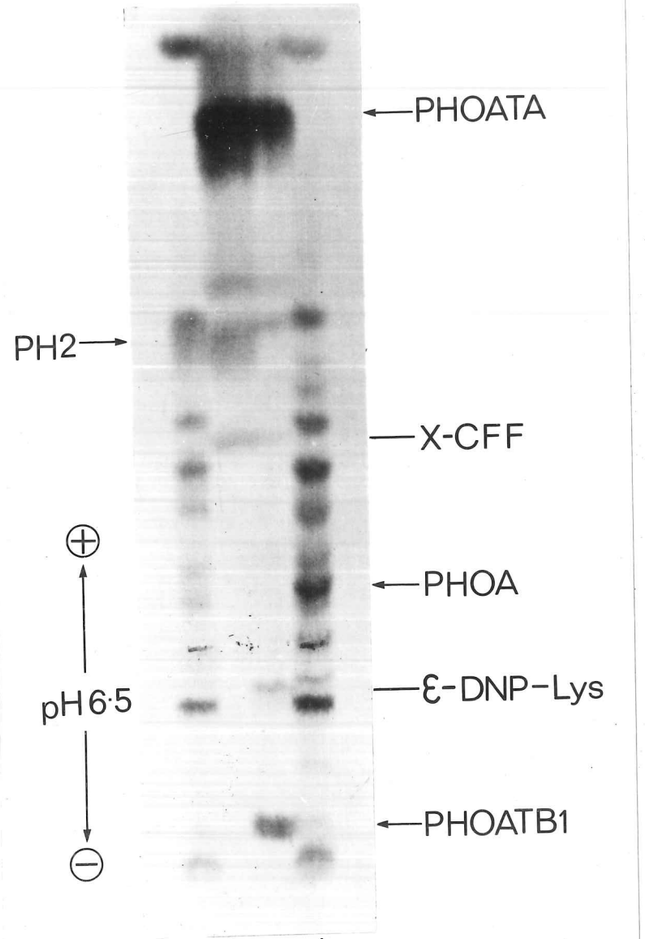


Fig. 6.14
Demonstration that the interchain disulphide bridge peptide of IF-2 H-chain (PH2) is a modified form of that in the wild-type (PHOA) by electrophoresis at pH 6.5 and radioautography of the ^{14}C-Lys, ^{14}C-Arg labelled material. a) Peptic digest of IF-2 IgG; b) Tryptic digest of PH2; c) Tryptic digest of PHOA; d) Peptic digest of P3K IgG. The major digestion product of PH2 is clearly identical to that of PHOA (PHOATA). This peptide has a very characteristic mobility and is known to contain cysteic acid residues derived from the interheavy chain disulphide bridges (Svasti & Milstein, 1972).

H-chain fingerprints of mutant and wild-type were largely similar,(Fig.6.12), again suggesting that the mutation involved an alteration in the structural gene coding for P3K H-chain and not the inactivation of that gene and the activation of another. There were a small number of differences however: the P3K peptides 4, 7, 12α were missing in the IF-2 fingerprint. Peptide 4 is Val-Asp-Lys-Lys (210-213), 7 is Ile-Val-Pro-Arg (214-217) and 12α is the peptide containing the interchain disulphide bridges (218-242). All these are from the "hinge" region of the H-chain. Peptides from the N- and C-terminal regions were common to P3K and IF-2.

6.4.2 Fingerprints of Peptic Peptides

In the fingerprint of a peptic digest of ^{35}S-Cys labelled IF-2 IgG (Fig.6.13), peptides C3 and C15b (from Cys III) and peptide PHOA which contains Cys IV and those Cys residues involved in interchain disulphide bridges were absent (Svasti & Milstein, 1972a). All other Cys-containing peptides were present in IF-2. A peptide unique to IF-2 (PH2) appeared to have replaced PHOA suggesting a modified, rather than missing, interchain peptide. This was confirmed by further digestion of PHOA and PH2 with trypsin resulting in a common acidic peptide containing the sequence around the interchain disulphide bridges (218-242 ; Fig. 6.14)

Thus the deletion is intramolecular and extends from residue 214, 215 or 216 towards the N-terminus and ends between residues 99 and 142. Subsequent work by K. Adegtubo has defined the position of this deletion more precisely(Fig. 6.1).

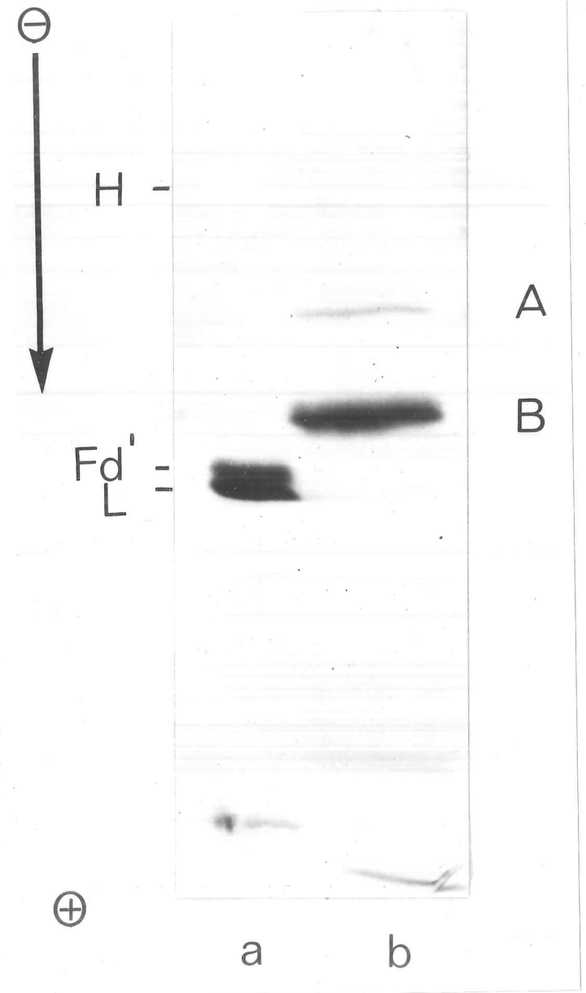


Fig. 6.15
Pepsin digest (pH 4.0) of mixed unlabelled P3K IgG and ^{14}C-Lys labelled mutant IgG. a) IF-1, b) IF-2.
Products analysed by SDS-polyacrylamide gel electrophoresis and radioautography of the dried gel. The position of the carrier P3K fragments (P3K H, Fd',L), visualised by staining the same gel with Coumassie blue dye, are indicated. Band A is undigested IF-2 H-chain. Band B has an approximate M.W. of 29,000

6.5 Enzymic Fragments of IF-1 and IF-2

In essence the deletion in IF-1 appears to be one of the C_{H3} domain and that in IF-2, of the C_{H1}. The remainder of the amino acid sequence of these proteins appears to be identical with the wild-type sequence. A knowledge of the way in which the remaining domains interact in the variant proteins could provide a valuable insight into the intramolecular interactions within Ig molecules. A simple way to explore these interactions is to examine the susceptibility of the Ig in its native conformation to mild proteolysis (Porter, 1959 ; Nisonoff et al, 1960).

^{14}C-labelled IF-1 and IF-2 IgG were each digested in the presence of added P3K IgG with pepsin and the digests examined by SDS-gel electrophoresis. Protein stain revealed the position of the L-chain and Fd' from P3K IgG (Svasti, 1972), whilst radioautography of the dried gel indicated the mobility of the fragments produced from the variant proteins. Fig.6.15 shows that in IF-1 the mutual protection of Fd and L-chain is apparently unaffected by the absence of C_{H3}. By contrast, in the absence of C_{H1} the L-chain of IF-2 is digested by the enzyme and the only fragment remaining is one of about 29,000 M.W. This is presumably an F_c fragment, an unusual finding for mouse IgG, and it suggests that in the absence of C_{H1} the Fc adopts an abnormal conformation.

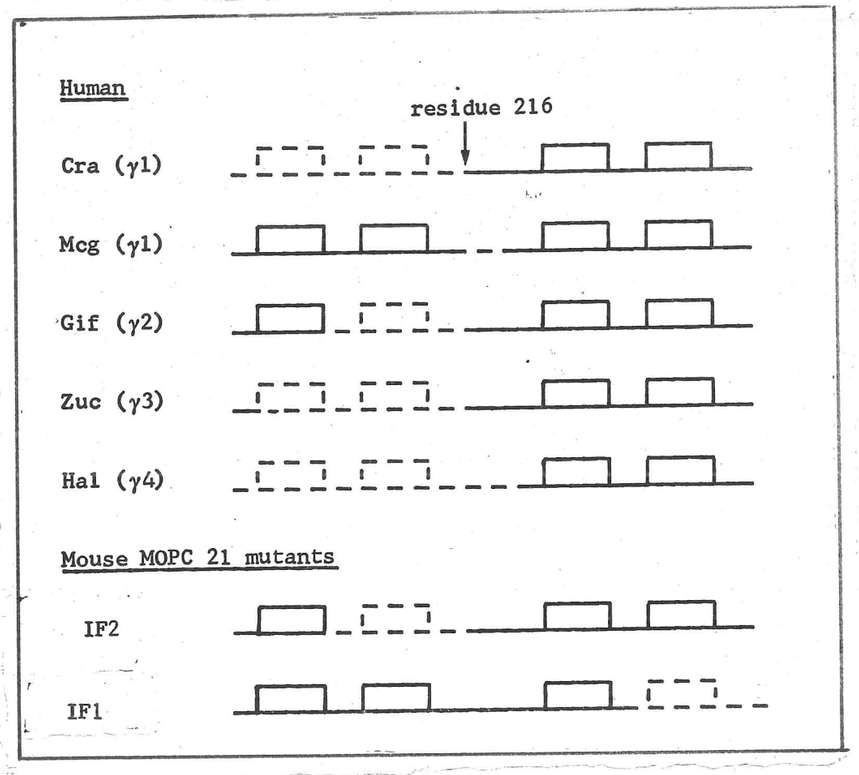


Fig. 6.16
A comparison of deletions in human heavy chains with those in the mouse mutants. The H-chain is shown diagramatically with its four intrachain disulphide bridges. Broken lines represent deleted regions of the molecule.

6.6 Discussion

The results presented demonstrate that mutants with alterations in their Ig structural genes can be isolated from tissue cultured cells. It is possible that the nature of the genetic events leading to the structural alterations of IF-1 and IF-2 is different. At this stage IF-1 could be explained by a point mutation, frame shift, or out of phase recombination giving rise to a chain termination triplet. IF-2, on the other hand, appears to be the result of a recombinational event leading to an internal deletion.

The C-terminus of IF-1 differs from that of the parental H-chain, which suggests that the deletion involves the C-terminal section. Proteolysis of a C-terminal portion seems excluded in view of the results obtained on intracellular and *in vitro* synthesised material. The differences between the carbohydrate of wild-type and IF-1 H-chains are likely to be secondary changes due to a failure to recognise the altered Fc correctly.

The mutation observed in IF-2 involves an interanl deletion of about 100 residues. Further analysis of the altered peptides will reveal the exact position of the deletion, but from the results obtained so far it is clear that it starts at or very near the junction of the V- and C-sections, and ends at or close to residue 216. A comparison of the IF-2 deletion with deletions observed in human heavy chain disease proteins is shown in Fig.6.16. Residue 216 of the four human γ-chain subclasses is glutamic acid (Frangione *et al*, 1969). The homologous position in mouse γ1 is valine (Svasti & Milstein, 1972a)

and it will be of interest to know whether it is at this same site that mutant IF-2 and wild-type H-chains regain homology.

In the case of human heavy chain disease the presumed parental protein has never been available. In contrast, the heavy chain of IF-2 can be compared with the MOPC21 heavy chain. Therefore one can ask whether the V-section of the deleted protein is identical to the parental one. The fingerprint evidence suggests that this is so, but confirmation by extensive sequence studies is required. If it proves to be the case it would strongly suggest that the deletion occurred after integration of V- and C-genes, providing further evidence that integration occurs at the DNA level (Cotton & Milstein, 1973). The fact that in mouse, as in humans, residue 216 appears to be a particular target in deletion events remains a puzzle. It is as if that part of the DNA is somehow similar to a signal for recombinational events, making it prone to this type of mistake. Whether this signal is related to those involved in V-C integration can only be a matter of speculation at this stage.

CHAPTER 7

Conclusion

Studies on immunoglobulins of rat (chapter 4) and mouse (chapters 5 and 6) origin have been described. The pattern of κ-chain variability in these species is markedly different from that of man which remains the simplest observed so far and that on which all hypotheses of antibody diversity are originally based. Yet chapter 4 shows that the rat Ig pattern, although equally complex to that of mouse, and despite the extreme evolutionary proximity of the two rodent species, is quite distinct. Only one set of sequences appeared to be common.

If the Ig V-genes expressed are selected at random from a similar pool in each species, then the number of germ-line genes is presumably very large. And yet the observation of sets of similar sequences within each species is the best evidence to date for a recent "expansion" (either in the germ-line or somatically) from a small number of genes (Milstein & Pink, 1970).

The invention of the sequenator has made possible the survey of a number of proteins in a fraction of the time it would have taken to determine the sequences conventionally. Surveys of myeloma protein sequences from other species are now required to provide more information on the evolution of Ig genes and any general hypothesis must explain the observations in all species studied.

It has become clear however that the accumulation of more protein sequence data will not, in the near future, resolve the problem of the origin of antibody diversity. Somatic mutation, as a plausible mechanism for the generation of diversity should be observed directly

in Ig structural genes. In chapter 5 an attempt to observe such mutation is described. The myeloma cell culture proved to be an excellent system. In culture the cells incorporate radioactive amino acids efficiently and the use of these labels to characterise the secreted protein, and fragments derived from it, extending the work of Svasti (1972), proved a rapid and efficient means of comparison. The P3K line was uniquely suitable in that Ig and mRNA are both amenable to study and completely or partially sequenced. The adaptation of variant cells to growth as solid tumours producing Ig in mice allowed study of non-radioactive protein to confirm and complement the results on the labelled material.

Variants were found and in chapter 6 the chemical analysis characterising these variants and showing that, in two cases, they result from structural gene mutations, is detailed.

The mutations involved deletions of whole domains or "pseudo-subunits" although the technique of isoelectric focusing is capable of detecting much less drastic alterations.

The relevance of these mutations to the role of somatic mutation in the generation of antibody diversity is not clear. Indeed any generalisation based on the characterisation of only two mutations is necessarily speculative (Milstein et al, 1974b). However the initial importance of the isolation of spontaneous somatic mutants in animal cells lies not in their immunological implications, but in the demonstration that they exist at all! Previous observations of somatic cell mutants had been made in circumstances under which it was impossible to localise the mutation to a structural gene rather than, say, a control gene or to an "auxiliary" gene producing (or controlling) for example a carbohydrate addition (Coffino et al, 1972; Tobari & Kojima, 1972). The wide range of mutation frequencies reported for eukaryotic cells may reflect a variety of mechanisms causing mutation. Using the system

described it is possible to determine the frequency of a particular type of mutational event.

The most frequent phenotypic variation was the cessation of IgG secretion (Table 5.1). It has already been shown that these are not a single class of variants (Milstein et al, 1974b). An understanding of the molecular defects in these cells may indicate the levels of control in Ig biosynthesis.

The variant proteins secreted by mutant cells have been investigated so far only at the level of primary structure. Their use as tools to investigate the folding, assembly and function of normal IgG has not yet been exploited, but if the variety as well as the number of mutants available continues to increase, a battery of variants could be studied.

The variants detected so far have all arisen spontaneously. Mutagens were avoided (as far as possible) so that the nature and frequency of variants should reflect the normal situation. It is premature to suggest that the different results obtained with (Preud'homme et al, 1973) and without (this study) mutagens are significant.

The emergence and survival of mutant cells in higher animals is complicacated by the simeltaneous and opposing trends of somatic mutation and selection (operating by altered cell division rates, immunological surveillance, loss of essential functions, etc.). In a system of continuous culture, selection is partially eliminated by the maintenance of a constant and non-interacting environment. It may be impossible to remove selective pressure entirely, but in the case of a myeloma cell culture, where the the variants detected are in a secreted protein that confers no apparent benefit to the cell which produced it, such pressures are presumed to be minimal.

It may be that a hypermutation mechanism exists in precursor cells but is inoperative in the differentiated plasma cell. The nature and frequency of variation in a less differentiated Ig-producing cell line

would provide an important comparison with the myeloma line data.

Finally the use of selective techniques to screen for variation in specific regions or functions of the Ig molecule would allow screening of much larger populations of cells and provide more accurate frequencies of less common mutations.

SUMMARY

Two approaches to evaluating the contribution of somatic mutation to antibody diversity have been made.

The first approach was to compare immunoglobulin sequences from a single species. A survey of amino-terminal sequences from 25 rat κ-chains was carried out using automatic protein sequencing techniques. With a modification to the standard procedure at least 30 residues of sequence were obtained in most runs.

Computer comparison of the rat sequences revealed a degree of complexity similar to that observed in mouse κ-chains by others. However individual mouse and rat sequences showed little homology.

Secondly a technique for the direct observation of mutation in immunoglobulin-producing cells is described. Cells grown in continuous culture were selected at random and cloned in soft agar. After development of the clones and the establishment of replicate cultures the products of each clone were examined by *in vivo* incorporation of ^{14}C-lysine and isoelectric focusing of the secreted material.

Clones which had ceased to secrete immunoglobulin and, at a lower frequency (10^{-5} - 10^{-6} /cell/generation), clones which secreted variant immunoglobulin were detected. Two of the latter clones (IF-1 and IF-2) were studied in detail.

Fingerprint analysis of radiolabelled protein suggested that IF-1 produced a Heavy chain lacking about 100 residues from the carboxy-terminus. Analysis by isoelectric focusing of the newly synthesised protein and characterisation of the mRNA coding for IF-1 *in vitro* confirmed that this deletion was due to mutation in the structural gene coding for the Heavy chain.

IF-2 also produced an unusually short Heavy chain. The deletion was shown to be internal and about 100 residues long. The amino-

terminal sequence and the half-cystine peptides of the Variable region were identical to those of the wild-type Heavy chain. A structural gene mutaion seems the most likely explanation for IF-2 also.

REFERENCES

Acher, R. & Crocker, C. (1952) Biochim. Biophys. Acta 9, 704-705.

Africa, B. & Carpenter, F.H. (1966) Biochem. Biophys. Res. Commun. 24, 113-119

Ambler, R.P. (1963) Biochem. J. 89, 341-349

Ambler, R.P. (1972) in Methods in Enzymology (Hirs, C.H.W. & Timasheff, S.N., eds.) vol.25, pp.143-154, Academic Press, New York and London

Anfinsen, C.B., Åqvist, S.E.G., Cooke, J.E. & Jönsson, B. (1959) J. Biol. Chem. 234, 1118-1123

Appella, E. & Ein, D. (1967) Proc. Nat. Acad. Sci. U.S. 57, 1449-1454

Askonas, B.A. (1961) Biochem. J. 79, 33-43

Askonas, B.A., Williamson, A.R. & Wright, B.E.G. (1970) Proc. Nat. Acad. Sci. U.S. 67, 1398-1403

Awdeh, Z.L. (1969) Science Tools 16, 42-43

Awdeh, Z.L., Williamson, A.R. & Askonas, B.A. (1968) Nature (London) 219, 66-67

Awdeh, Z.L., Williamson, A.R. & Askonas, B.A. (1970) Biochem. J. 116, 241-248

Baernstein, H.D. (1936) J. Biol. Chem. 115, 33-36

Baglioni, C., Zonta, L.A., Cioli, D. & Carbonara, A. (1966) Science 152, 1517-1519

Benditt, E.P., Eriksen, N., Hermodson, M.A. & Ericsson, L.H. (1971) FEBS Lett. 19, 169-173

Bennich, H. & Johansson, S.G.O. (1971) Adv. Immunol. 13, 1-55

Bennich, H., Milstein, C. & Secher, D.S. (1973) FEBS Lett. 33, 49-53

Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. & Petersen, U. (1971) Hoppe-Seyler's Z. Physiol. Chem. 352, 1730-1732

Braunitzer, G., Chen, R., Schrank, B. & Stangl, A. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 867-878

Bray, G.A. (1960) Anal. Biochem. 1, 279-285

Brenner, S. & Milstein, C. (1966) Nature (London) 211, 242-243

Breslow, R.E. & Goldsby, R.A. (1969) Exp. Cell Res. 55, 339-346

Bretscher, M. (1971) Nature (London), New Biol. 231, 229-232

Bridgen, J. & Secher, D.S. (1973) FEBS Lett. 29, 55-57

Britten, R.J. & Kohne, D.E. (1967) Carnegie Inst. Yearbook 66, 73-88

Brown, D.D. (1972) in "Molecular Genetics and Developmental Biology" (Sussman, M. ed.) pp.101-125, Prentice Hall, Englewood Cliffs, N.J.

Brownlee, G.G. & Sanger, F. (1969) Eur. J. Biochem. 11, 395-399

Brownlee, G.G., Harrison, T.M., Mathews , M.B. & Milstein, C. (1972) FEBS Lett. 23, 244-248

Brownlee, G.G., Cartwright, E.M., Cowan, N.J., Jarvis, J.M. & Milstein, C. (1973) Nature (London), New Biol. 244, 236-240

Burnet, F.M. (1959) "The Clonal Selection Theory of Acquired Immunity Vanderbilt Univ. Press, Nashville, Tenn.

Burnet, F.M. (1969) "Cellular Immunology" University Press, Cambridge

Callan, H.G. (1967) J. Cell Sci. 2, 1-7

Capra, J.D. & Kunkel, H.G. (1970) Proc. Nat. Acad. Sci. U.S. 67, 87-92

Cassidy, J.T., Jourdian, G.W. & Roseman, S. (1965) J. Biol. Chem. 240, 3501-3506

Chu, E.H.Y., Brimer, P., Jacobson, K.B. & Merrian, E.V. (1969) Genetics 62, 359-377

Coffino, P. & Scharff, M.D. (1971) Proc. Nat. Acad. Sci. U.S. 68, 219-223

Coffino, P., Baumal, R., Laskow, R. and Scharff, M.D. (1972) J. Cell Physiol. 79, 429-440

Cohn, M. (1968) in "Nucleic Acids in Immunology (Plescia, D.J. & Brown, W. eds.) pp.671-715 Springer-Verlag, New York

Cohn, M. (1971) Ann. N.Y. Acad. Sci. 190, 529-583

Cotton, R.G.H. & Crawford, I.P. (1972) J. Biol. Chem. 247, 1883-1891

Cotton, R.G.H. & Milstein, C. (1973) Nature (London) 244, 42-43

Cotton, R.G.H., Secher, D.S. & Milstein, C (1973) Eur. J. Immunol. 3, 135-140

Cowan, N.J. & Milstein, C. (1973) Eur. J. Biochem. 36, 1-7

Cowan, N.J. & Milstein, C. (1974) J. Mol. Biol. 82, 469-481

Cowan, N.J., Secher, D.S., Cotton, R.G.H. & Milstein, C. (1973) FEBS Lett. 30, 343-346

Crick, F.H.C. (1958) Symp. Soc. Exp. Biol. 12, 138-163

Crowshaw, K., Jessup, S.J. & Ramwell, R.W. (1967) Biochem. J. 103, 79-85

Cunningham, B.A., Gottlieb, P.D., Pflumm, M.N. & Edelman, G.M. (1971) Progr. Immunol. 1, 3-24

Dayhoff, M.O. (ed.) (1972) "Atlas of Protein Sequence and Structure" vo. 5, National Biomedical Research Foundation, Silver Spring

Dreyer, W.J. & Bennett, J.C. (1965) Proc. Nat. Acad. Sci. U.S. 54, 864-869

Dreyer, W.J., Gray, W.R. & Hood, L. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 353-367

Eagle, H. (1959) Science 130, 432-437

Easley, C. (1965) Biochim. Biophys. Acta 107, 386-388

Easley, C.W., Zegers, B.J.M. & de Vijlder, M. (1969) Biochim. Biophys. Acta 175, 211-213

Edelman, G.M. & Gally, J.A. (1968) Brookhaven Symp. Biol. 21, 328-343

Edelman, G.M., Cunningham, B.A., Gall, W.E., Gottlieb, P.D., Rutishauser, U. & Waxdal, M.J. (1969) Proc. Nat. Acad. Sci. U.S. 63, 78-85

Edman, P. (1950) Acta Chem. Scand. 4, 283-293

Edman, P. (1970) in "Protein Sequence Determination" (Needleman, S.B., ed.) pp.211-255, Springer-Verlag, Berlin

Edman, P. & Begg, G. (1967) Eur. J. Biochem. 1, 80-91

Edwards, A.W.F. & Cavalli-Sforza, L.L. (1963) Annals Hum. Genet. 27 104-105

Edwards, A.W.F. & Cavalli-Sforza, L.L. (1964) in "Phenetic and Phylogenetic Classification" pp.67-76 Systematics Association, London

Ein, D. & Hood, L. (1968) Fed. Proc. 27, 559

Eisen, H.N. (1971) Progr. Immunol. 1, 243-251

Eulitz, M. & Hilschmann, N. (1974) Hoppe-Seyler's Z. Physiol. Chem. 355, 842-866

Feinstein, A. & Munn, E.A. (1969) Nature (London) 224, 1307-1309

Feinstein, A., Gell, P.G.H. & Kelus, A.S. (1963) Nature (London) 200, 653-654

Fisher, C.E. & Press, E.M. (1974) Biochem. J. 139, 135-149

Fitch, W.M. (1966) J. Mol. Biol. 16, 9-16

Fitch, W.M. & Margoliash, E. (1967) Science 155, 279-284

Fleet, G.W.J., Porter, R.R. & Knowles, J.R. (1969) Nature (London) 224, 511-512

Fleet, G.W.J., Knowles, J.R. & Porter, R.R. (1972) Biochem. J. 128, 499-508

Franék, F. (1971) Eur. J. Biochem. 19, 176-183

Frangione, B. & Franklin, E.C. (1973) Semin. Hemat. 10, 53-64

Frangione, B., Milstein, C. & Pink, J.R.L. (1969) Nature (London) 221, 145-148

Frank, G. & Strubert, W. (1973) Chromatographia 6, 522-524

Franklin, E.C. & Frangione, B. (1971) Proc. Nat. Acad. Sci. U.S. 68, 187-191

Fraser, A.S. & Short, B.F. (1958) Aust. J. Biol. Sci. 11, 200-208

Gally, J.A. & Edelman, G.M. (1970) Nature (London) 227, 341-348

Gally, J.A. (1973) in "The Antigens" (Sela, M. ed.) vol.1, pp.161-298, Academic Press, London

Ghanem, N.A. & Westermark, T. (1960) J. Amer. Chem. Soc. 82, 4432-4433

Gosztonyi, T. & Walde, N. (1966) J. Labelled Comp. 2, 155-173

Gottschalk, A. (1966) "Glycoproteins" pp.15-16, Elsevier, London

Graffeo, A.P., Haag, A., Karger, B.L. (1973) Anal. Lett. 6, 505-511

Gray, W.R. (1968) Nature (London) 220, 1300-1304

Gray, W.R. (1972) in "Methods in Enzymology" (Hirs, C.H.W. & Timasheff, N. ed.) vol.26, pp.333-344, Academic Press, London

Gray, W.R. & Hartley, B.S. (1963) Biochem. J. 89, 59P

Gray, W.R., Dreyer, W.J. & Hood, L. (1967) Science 155, 465-467

Heilmann, J., Barrollier, J. & Watzke, E. (1957) Hoppe-Seyler's Z. Physiol. Chem. 309, 219-220

Hembree, W.C., Ehrenkaufer, R.E., Lieberman, S. & Wolf, A.P. (1973) J. Biol. Chem. 248, 5532-5540

Hermodson, M.A., Ericsson, L.H., Titani, K., Neurath, H. & Walsh, K.A. (1972) Biochemistry 11, 4493-4502

Hess, M., Hilschmann, N., Rivat, L., Rivat, C. & Roparts, C. (1971) Nature (London), New Biol. 234, 58-61

Hill, R.L., Delaney, R., Fellows, R.E. & Lebovitz, H.E. (1966) Proc. Nat. Acad. Scie. U.S. 56, 1762-1769

Hilschmann, N. & Craig, L.C. (1965) Proc. Nat. Acad. Sci. U.S. 53, 1403-1409

Hilschmann, N., Barnikol, H.U., Hess, M., Langer, B., Ponstingl, H., Steinmetz-Kayne, M., Suter, L. & Watanabe, S. (1969) in "Gamma Globulins - Structure and Biosynthesis" (Franek F. & Shugar, D. eds.) FEBS Symposium vol.15, pp.57-74, Academic Press, London

Hood, L. (1973) Stadler Symp. 5, 73-142

Hood, L. & Talmage, D.W. (1970) in "Developmental Aspects of Antibody Formation and Structure" (Sterzl, J. & Riha, I. eds.) pp.935-962, Academic Press, New York

Hood, L., Gray, W.R., Sanders, B.G. & Dreyer, W.J. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 133-146

Hood, L., McKean, D., Farnsworth, V. & Potter, M. (1973) Biochemistry 12, 741-771

Horibata, K. & Harris, A.W. (1970) Exp. Cell Res. 60, 61-77

Imamura, T., Baldwin, T.O. & Riggs, A. (1972) J. Biol. Chem. 247, 2785-2797

Inglis, A.S., Nichols, P.W. & Roxburgh, C.M. (1971) Aust, J. Biol. Sci. 24, 1235-1240

Ingram, V.M. (1957) Nature (London) 180, 326-328

Inman, J.K. (1974) in "The Immune System: Genes, Receptors, Signals" (Sercarz, E.E., Williamson, A.R. & Fox, C.F. eds.) pp.37-52, Academic Press, New York

Jeppesen, P.G.N., Nichols, J.L., Sanger, F. & Barrell, B.G. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 13-19

Jerne, N.K. (1955) Proc. Nat. Acad. Sci. U.S. 41, 849-857

Jerne, N.K., (1971) Eur. J. Immunol. 1, 1-9

Kabat, E.A. & Wu, T.T. (1972) Proc. Nat. Acad. Sci. U.S. 69, 960-964

Kehoe, J.M. & Capra, J.D. (1971) Proc. Nat. Acad. Sci. U.S. 68, 2019-2021

Kehoe, J.M. & Fougereau, M. (1969) Nature (London) 224, 1212-1213

Kelus, A.S. & Gell, P.G.H. (1967) Progr. Allergy 11, 141-184

Koch, R.E. (1971) Proc. Nat. Acad. Sci. U.S. 68, 773-776

Kolb, E., Harris, J.I. & Bridgen, J. (1974) Biochem. J. 137, 185-197

Kölsch, E. (1967) J. Immunol. 98, 854-859

Koo, P.H. & Cebra, J. (1974) Biochemistry 13, 184-195

Krause, R.M. (1970) Adv. Immunol. 12, 1-56

Kreth, H.W. & Williamson, A.R. (1973) Eur. J. Immunol. 3, 141-147

Kunkel, H.G., Natvig, J.B. & Joslin, F.G. (1969) Proc. Nat. Acad. Sci. U.S. 62, 144-149

Laemmli, U.K. & Faüre, M. (1973) J. Mol. Biol. 80, 575-599

Landsteiner, K. (1945) "The Specificity of Serological Reactions", Harvard

Laursen, R.A. (1969) Biochem. Biophys. Res. Commun. 37, 663-667

Levin, A.S., Fudenberg, H.H., Hopper, J.E., Wilson, S.K. & Nisonoff, A. (1971) Proc. Nat. Acad. Sci. U.S. 68, 169-171

Li, S.L. & Riggs, A. (1970) J. Biol. Chem. 245, 6149-6169

Lockhard, R. & Lingrel, J.B. (1969) Biochem. Biophys. Res. Commun. 37, 204-212

Lynn, J.D. & Bennett, J.C. (1972) Anal. Biochem. 45, 498-509

McKean, D., Potter, M. & Hood, L. (1973) Biochemistry 12, 749-759

McKean, D.J., Peters, E.H., Waldby, J.I. & Smithies, O. (1974) in press Biochemistry

Mach, B., Faust, C. & Vassalli, P. (1973) Proc. Nat. Acad. Sci. U.S. 70, 451-455

Mage, R. (1971) Progr. Immunol. 1, 47-59

Mage, R., Lieberman, R., Potter, M. & Terry, W.D. (1973) in "The Antigens" (Sela, M. ed.) vol.1, pp.299-376, Academic Press, New York

Milstein, C. (1964) J. Mol. Biol. 9, 836-838

Milstein, C. (1966a) Nature (London) 209, 370-373

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

Milstein, C. (1967) Nature (London) 216, 330-332

Milstein, C. & Munro, A.J. (1970) Ann. Rev. Microbiol. 24, 335-358

Milstein, C. & Munro, A.J. (1973) in "Defence and Recognition" (Porter, R.R. ed.) pp.199-228, Butterworth, London

Milstein, C. & Pink, J.R.L. (1970) Progr. Biophys. Mol. Biol. 21, 209-263

Milstein, C. & Svasti, J. (1971) Progr. Immunol. 1, 33-45

Milstein, C., Milstein, C.P. & Feinstein, A. (1969) Nature (London) 221, 151-154

Milstein, C., Brownlee, G.G., Harrison, T.M. & Mathews, M.B. (1972) Nature (London), New Biol. 239, 117-120

Milstein, C., Secher, D.S., Cowan, N.J., Harrison, T.M., Cotton, R.G.H. & Brownlee, G.G. (1974a) in "Cellular Selection and Regulation in the Immune Response" (Edelman, G.M. ed.) pp.245-264, Raven Press, New York

Milstein, C., Adetugbo, K., Cowan, N.J. & Secher, D.S. (1974b) in press Proc. 2nd Int. Congr. Immunol.

Mole, L.E., Jackson, S.A., Porter, R.R. & Wilkinson, J.M. (1971) Biochem. J. 124, 301-318

Mushinski, J.F. (1971) J. Immunol. 106, 41-50

Natori, S. & Garen, A. (1970) J. Mol. Biol. 49, 577-588

Niall, H.D. (1971) J. Agr. Food Chem. 19, 638-644

Niall, H.D. (1973) in "Methods in Enzymology" (Hirs, C.H.W. & Timasheff, S.N. eds.) vol.27, pp.942-1010, Academic Press, London

Niall, H.D., Keutmann, H.T., Copp, D.H. & Potts, J.T., Jr. (1969) Proc. Nat. Acad. Sci. U.S. 64, 771-778

Niall, H.D., Keutmann, H., Sauer, R., Hogan, M., Dawson, B., Aarbach, G. & Potts, J.,Jr. (1970) Hoppe Seyler's Z. Physiol. Chem. 351, 1586-1588

Niall, H.D., Jacobs, J.W., van Rietshoton, J. & Tregear, G.W. (1974) FEBS Lett. 41, 62-64

Nisonoff, A., Wissler, F.C., Lipman, L.N. & Woernley, D.L. (1960) Arch. Biochem. Biophys. 89, 230-244

Nossal, G.J.V., Szenberg, A., Ada, G.L. & Austin, C.M. (1964) J.Exp. Med. 119, 485-503

Novotny, J. (1973) J. Theor. Biol. 41, 171-180

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

Oudin, J. & Michel, M. (1969) J. Exp. Med. 130, 619-642

Paul, J. (1970) "Cell and Tissue Culture" p.239 E. & S. Livingstone, London

Pernis, B., Chiappino, G., Kelus, A.S. & Gell, P.G.H. (1965) J. Exp. Med. 122, 853-875

Pink, J.R.L., Buttery, S.H., de Vries, G.M. & Milstein, C. (1970) Biochem. J. 117, 33-47

Pink, J.R.L., McNally, M.P., Wang, A.C. & Fudenberg, H.H. (1972) Immunochem. 9, 84-86

Pisano, J.J. & Bronzert, T.J. (1969) J. Biol. Chem. 244, 5597-5607

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P. & Saul, F. (1973) Proc. Nat. Acad. Sci. U.S. 70, 3305-3310

Porter, R.R. (1959) Biochem. J. 73, 119-126

Porter, R.R. (1973) Science 180, 713-716

Potter, M. (1972) Physiol. Rev. 52, 631-719

Potter, M., Dreyer, W.J., Kuff, E.L. & McIntire, K.R. (1964) J. Mol. Biol. 8, 814-822

Potter, M. & Lieberman, R. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 187-202

Premkumar, E., Shoyab, M. & Williamson, A.R. (1974) Proc. Nat. Acad. Sci. U.S. 71, 99-103

Preud'homme, J.-L., Buxbaum, J. & Scharff, M.D. (1973) Nature (London) 245, 320-322

Quattrochi, R., Cioli, C. & Baglioni, C. (1969) J. Exp. Med. 130, 401-415

Querinjean, P., Bazin, H., Beckers, A., Deckers, C., Heremans, J.F. & Milstein, C. (1972) Eur. J. Biochem. 31, 354-359

Ramshaw, J.A.M., Scawen, M.D., Bailey, C.J. & Boulter, D. (1974) Biochem. J. 139, 583-592

Sanger, F. (1958) Proc. 4th Int. Cong. Biochem., I.U.B. Symp. vol.10, p.106, Pergamon Press, London

Sanger, F. & Thompson, E.O.P. (1963) Biochim. Biophys. Acta 71, 468-471

Sanger, F. & Tuppy, H. (1951) Biochem. J. 49, 463-481

Sarabhai, A.S., Stretton, A.O.W., Brenner, S. & Bolle, A. (1964) Nature (London) 201, 13-17

Schechter, I. (1973) Proc. Nat. Acad. Sci. U.S. 70, 2256-2260

Schiffer, M., Girling, R.L., Ely, K.R. & Edmunson, A.B. (1973) Biochemistry 12, 4620-4631

Schwartz, J. & Edelman, G.M. (1963) J. Exp. Med. 118, 41-53

Schwert, G.W. & Takenaka, Y. (1955) Biochim. Biophys. Acta 16, 570-575

Secher, D.S., Cotton, R.G.H. & Milstein, C. (1973) FEBS Lett. 37, 311-316

Shaw, C.R. (1965) Science 149, 936-943

Singer, S.J. & Doolittle, R.F. (1966) Science 153, 13-25

Smith, G.P. (1973) "The Variation and Adaptive Expression of Antibodies" Harvard University Press, Cambridge, USA

Smith, G., Hood, L. & Fitch, W. (1971) Ann. Rev. Biochem. 40, 969-1012

Smith, J.D., Freeman, G. Vogt, M. & Dulbecco, R. (1960) Virology 12 185-196

Smith, M., Stavnezer, J., Huang, R.C., Gurdon, J.B. & Lane, C.D. (1973) J. Mol. Biol. 80, 553-557

Smithies, O. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 161-168

Smithies, O. (1970) Science 169, 882-883

Smithies, O., Gibson, D., Fanning, E.M., Goodfleisch, R.M., Gilman, J.G. & Ballantyne, D.L. (1971) Biochem. 10, 4912-4921

Spencer, E.M. & King, T.P. (1971) J. Biol. Chem. 246, 201-208

Stavnezer, J. & Huang, R.C. (1971) Nature (London), New Biol. 230, 172-176

Stevens, R.H. & Williamson, A.R. (1972) Nature (London) 239, 143-146

Stevenson, G.T. & Mole, L.E. (1974) Biochem. J. 139, 369-374

Svasti, J. (1972) Ph.D. Thesis, University of Cambridge

Svasti, J. & Milstein, C. (1972a) Biochem. J. 126, 837-850

Svasti, J. & Milstein, C. (1972b) Biochem. J. 128, 427-444

Swan, D., Aviv, H. & Leder, P. (1972) Proc. Nat. Acad. Sci. U.S. 69, 1967-1971

Schulenberg, E.P., Simms, E.S., Lynch, R.G., Bradshaw, R.A. & Eisen, H.N. (1971) Proc. Nat. Acad. Sci. U.S. 68, 2623-2626

Talmage, D.W. (1959) Science 129, 1643-1648

Terhorst, C., Müller, W., Laursen, R. & Wittmann-Liebold, B. (1973) Eur. J. Biochem. 34, 138-152

Titani, K. & Putnam, F.W. (1965) Science 147, 1304-1305

Tobari, Y.N. & Kojima, K. (1972) Genetics 70, 397-403

Todd, C.W. (1963) Biochem. Biophys. Res. Commun. 11, 170-175

Todd, C.W. & Inman, F.P. (1967) Immunochem. 4, 407-417

Tonegawa, L.S., Bernadini, A., Weimann, B.J. & Steinberg, C. (1974) FEBS Lett. 40, 92-96

Turner, M.W. & Bennich, H. (1968) 107, 171-178

Valentine, R.C. & Green, M.N. (1967) J. Mol. Biol. 27, 615-617

van Orden, H.O. & Carpenter, F.H. (1964) Biochem. Biophys. Res. Commun. 14, 399-403

Varga, J.M., Konigsberg, W.H. & Richards, F.F. (1973) Proc. Nat. Acad. Sci. U.S. 70, 3269-3274

Waley, S.G. & Watson, J. (1953) Biochem. J. 55, 328-337

Wang, A.C., Wilson, S.K., Hopper, J.E., Fudenberg, H.H. & Nisonoff, A. (1970) Proc. Nat. Acad. Sci. U.S. 66, 337-343

Wang, A.C., Fudenberg, H.H. & Pink, J.R.L. (1971) Proc. Nat. Acad. Sci. U.S. 68, 1143-1146

Waterfield, M.D., Corbett, C. & Haber, E. (1970) Anal. Biochem. 38, 475-492

Weygand, F. & Obermeier, R. (1971) Eur. J. Biochem. 20, 72-80

Whitehouse, H.L.K. (1973) Nature (London) 245, 295-298

Williamson, A.R. (1972) Biochem. J. 130, 325-333

Wittmann-Liebold, B. (1973a)FEBS Lett. 36, 247-249

Wittmann-Liebold, B. (1973b) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415-1431

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

Wu, T.T. & Kabat, E.A. (1970) J. Exp. Med. 132, 211-250

Zimmerman, C.L., Pisano, J.J. & Appella, E. (1973) Biochem. Biophys. Res. Commun. 55, 1220-1224

ABBREVIATIONS

The following non-standard abbreviations have been used in the text:

hr.	hour
min.	minute
sec.	second (time)
DEAE-	O-(diethylaminoethyl)-
DNP-	2,4-dinitrophenyl
RNase	ribonuclease
uv	ultraviolet
P.T.F.E.	polytetrafluorethylene ("Teflon")
PTH-	phenylthiohydantoin
TLC	thin layer chromatography
GLC	gas liquid chromatography
mT	millitorr
i.d.	internal diameter
CMCys	carboxymethyl cysteine
PITC	phenyl isothiocyanate
HFBA	heptafluorobutyric acid
PCA	pyrrolidone carboxylic acid
cpm	counts/min
M.W.	molecular weight
$\underline{m}$	electrophoretic mobility at pH 6.5
Tris	tris(hydroxymethyl)aminomethane
XCFF	xylene cyanol FF
SDS	sodium dodecyl sulphate
ψ-	pseudo-

The one-letter code for amino acids has also been used:

A	Ala	M	Met
B	Asx	N	Asn
C	Cys	P	Pro
D	Asp	Q	Gln
E	Glu	R	Arg
F	Phe	S	Ser
G	Gly	T	Thr
H	His	V	Val
I	Ile	W	Trp
K	Lys	Y	Tyr
L	Leu	Z	Glx

For the standard nomenclature used for immunoglobulins see Eur. J. Immunol. (1973) 3, 62.

For all other standard abbreviations see Biochem. J. (1973) 31, 1-20.