*Biochem. J.* (1981) **195**, 15–30
*Printed in Great Britain*



# Rat brain Thy-1 glycoprotein

## The amino acid sequence, disulphide bonds and an unusual hydrophobic region

David G. CAMPBELL,*† Jean GAGNON,‡ Kenneth B. M. REID‡ and Alan F. WILLIAMS*§
*M.R.C. Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, U.K. and ‡M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(*Received 20 June 1980/Accepted 4 December 1980*)

The full sequence of the Thy-1 membrane glycoprotein of rat brain is reported. The sequence was determined from tryptic and V-8 proteinase peptides and consisted of 111 amino acids. The amino terminus was blocked and consisted of a pyroglutamic acid residue. The molecule contained two disulphide bonds, namely Cys-9–Cys-111 and Cys-19–Cys-85. Three *N*-linked amino sugars were located at Asn-23, Asn-74 and Asn-98. In each case the sequence on the *C*-terminal side of the attachment point was Asn-Xaa-Thr as would be expected for *N*-linkage. The *C*-terminal peptides were unusual, in that they were either obtained in a highly aggregated form, or could only be purified after binding to Brij 96 micelles. Thus they appeared to have hydrophobic properties, yet did not contain any extended sequence of hydrophobic amino acids. Other unusual features of the *C*-terminal peptides were the presence of unidentified ninhydrin-positive material and of glucosamine and galactosamine. The *C*-terminal residue has not been directly identified but Cys-111 is the last conventional amino acid. It is suggested that the hydrophobic properties of the *C*-terminal peptides may be due to the linkage of lipid. The sequence of the Thy-1 glycoprotein showed homologies with immunoglobulin domains. This relationship is examined in detail in the paper following [Cohen *et al.* (1981) *Biochem. J.* **193**, 000–000].

The Thy-1 (Θ) antigen was first identified in the mouse as a cell surface alloantigen of thymus and brain, with two allotypic forms called Thy-1.1 (Θ-AKR) and Thy-1.2 (Θ-$C_3H$) (Reif & Allen, 1964). The Thy-1.1 antigenic determinant was also found in rat thymus and brain (Douglas, 1972) but the Thy-1.2 determinant has not been detected in the rat. The mouse and rat Thy-1.1 determinants are similar but not identical, since the affinity of cross-reacting antibody differs by 10-fold between the two species (Mason & Williams, 1980). Other antigenic determinants on the Thy-1 molecule can be recognized by rabbit antiserum, and these include species-specific and cross-reacting determinants (reviewed in Williams *et al.*, 1976). Thy-1 antigens have also been identified in the dog and in humans by cross-reactions of rabbit antibodies to human or dog brain with rat Thy-1 (Arndt *et al.*, 1977; Dalchau & Fabre, 1979).

The tissue distribution of the Thy-1 antigen follows unusual patterns within a species and shows surprising differences between species. In all species studied the molecule is a major constituent of brain-cell membrane and is predominantly found on neuronal cells (Barclay & Hydén, 1978; Fields *et al.*, 1978) although some glial cells are also Thy-1-antigen positive (Pruss, 1979). In lymphoid tissues Thy-1 antigen is likely to be the most abundant cell-surface molecule of mouse and rat thymocytes (Acton *et al.*, 1974; Mason & Williams, 1980) but is found in reduced amounts in dog thymocytes and not at all in human thymocytes (Arndt *et al.*, 1977; Dalchau & Fabre, 1979). In the mouse, Thy-1 antigen is absent from stem cells and has been widely used as a marker for T lymphocytes (Raff,

Abbreviations used: SDS, sodium dodecyl sulphate; Xaa, unspecified amino acid residue; $iPr_2P$-F, di-isopropyl fluorophosphate; IgG, immunoglobulin G; Tos-Phe-$CH_2Cl$ ('TPCK'), 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one; dansyl, 5-dimethylamino-naphthalene-1-sulphonyl; CMCys, *S*-carboxymethyl-cysteine.

† Present address: The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, U.S.A.

§ To whom reprint requests should be addressed.



0306-3275/81/040015-16$01.50/1 

1971). In contrast to this, most rat T lymphocytes lack Thy-1 antigen (Acton *et al.*, 1974) and the molecule is found on a sub-set of bone-marrow cells (Williams, 1976), which include immature B lymphocytes and haemopoietic stem cells (Hunt *et al.*, 1977; Thierfelder, 1977; Goldschneider *et al.*, 1978). In rodents some fibroblasts (Stern, 1973), epidermal cells (Scheid *et al.*, 1972), breast cells (Lennon *et al.*, 1978) and muscle cells (Lesley & Lennon, 1977) also display Thy-1 antigen and the antigen is clearly detectable in post-capillary venules and some small capillaries in the rat (Ritter & Morris, 1980). In many of the above tissues the levels of Thy-1 expression undergo dramatic changes during differentiation.

All the Thy-1 antigenic determinants from thymus and brain of rats and mice are expressed in glycoproteins with an apparent molecular weight on polyacrylamide gels in SDS of about 25000, and the molecules have been purified from both tissues in both species (Barclay *et al.*, 1975; Letarte-Muirhead *et al.*, 1975; Trowbridge *et al.*, 1975; Zwerner *et al.*, 1977; Letarte & Meghji, 1978). The correct molecular weights, determined by sedimentation equilibrium measurements, are 17500 for brain Thy-1 and 18700 for thymocyte Thy-1, and in each case the molecular weight of the polypeptide is 12500 (Kuchel *et al.*, 1978). The amino acid compositions of Thy-1 glycoprotein from brain and thymus are very similar (Barclay *et al.*, 1976; McClain *et al.*, 1978) and the molecules from both sources share identical antigenic determinants, which appear to be protein-based (Williams *et al.*, 1976). In contrast to this, the carbohydrate composition obtained for brain Thy-1 glycoprotein differs from that found for the molecule from thymocytes. This accounts for the difference in molecular weight (Barclay *et al.*, 1976; McClain *et al.*, 1978).

Both thymocyte and brain Thy-1 glycoproteins have the properties of molecules that bind directly to the plasma membrane. Rat Thy-1 glycoprotein can be solubilized as a monomer in deoxycholate and Brij 96 and binds one micelle of deoxycholate/molecule. If the detergent is removed, the molecule self-associates to form an oligomer which is not dissociated by 4M-guanidinium chloride (Kuchel *et al.*, 1978). Mouse thymocyte Thy-1 glycoprotein is labelled in membranes by a photoactivatable reagent which partitions into lipid (Owen *et al.*, 1980). These properties suggest that the molecule has a hydrophobic portion in its structure and this is presumably responsible for integration into membrane since there has never been any indication of Thy-1 glycoprotein binding to another molecule in immunoprecipitation studies (Trowbridge *et al.*, 1975; Ledbetter & Herzenberg, 1979).

The determination of the sequence of rat brain Thy-1 glycoprotein was undertaken to elucidate further the properties of the molecule, particularly with respect to carbohydrate attachment sites and the nature of the hydrophobic region. Unexpectedly, sequence homologies with immunoglobulin were found.

## Experimental

### *Materials*

Wistar rats were obtained from the Oxfordshire Laboratory Animal Colonies, Bicester, U.K. and from the M.R.C. Radiobiology Unit, Harwell, Oxon., U.K.

Trypsin (treated with Tos-Phe-$CH_2Cl$) was from Worthington Biochemical Corp., Freehold, NJ, U.S.A. Pyroglutamate aminopeptidase (EC 3.4.11.8) was from Boehringer Corp., Lewes, Sussex, U.K. Pronase was from Kaken Chemical Co., Tokyo, Japan. V-8 proteinase from *Staphylococcus aureus* V8 was from Miles Laboratories, Stoke Poges, Slough, Bucks., U.K. Pepsin (twice crystallized) was from Sigma (London) Chemical Co., Poole, Dorset, U.K.

Bio-Gel P-10 (200–400 mesh) and Bio-Gel P-6 (200–400 mesh) were from Bio-Rad Labs., Bromley, Kent, U.K. Sephadex G-50 (fine grade) and Sephacryl S-200 were from Pharmacia Fine Chemicals, Hounslow, Middlesex, U.K.

Iodo[2-$^{14}$C]acetic acid (57mCi/mmol) was obtained from The Radiochemical Centre, Amersham, Bucks., U.K. Sodium deoxycholate, CNBr and 2,5-diphenyloxazole (PPO) were from Koch-Light Laboratories, Colnbrook, Bucks., U.K. Succinic anhydride was from M & B Laboratories, Dagenham, Essex, U.K. Bio-Fluor was from New England Nuclear. Quadrol (*NNN'N'*-tetrakis-(2-hydroxypropyl)ethylenediamine trifluoroacetate) was from Beckman Instruments, High Wycombe, Bucks., U.K. Polybrene (1,5-dimethyl-1,5-diazaundecamethylene polymethobromide) was from Aldrich Chemical Co., Gillingham, Dorset, U.K. All other chemicals used in automated sequencing were obtained from Rathburn Chemicals, Walkerburn, Peebleshire, Scotland, U.K. *p*-Toluenesulphonic acid was from Pierce Chemical Co., Rockford, IL, U.S.A.

Polyamide thin-layer sheets were from BDH Chemicals, Poole, Dorset, U.K.

### *Purification of rat brain Thy-1 glycoprotein*

The method of Barclay *et al.* (1975) was initially used but was subsequently modified to include a chloroform/methanol extraction step (suggested by Dr. J. Reynolds, Durham University, Durham, NC, U.S.A.) and affinity chromatography with a monoclonal antibody column. In this method 200g wet wt. of rat brain was homogenized in a Waring blender after adding 300ml of cold 0.16M-NaCl/

1mM-EDTA/50mM-Tris/HCl, pH8.0. The homogenate was mixed with 3.6 litres of chloroform/methanol (2:1, v/v) and vigorously stirred with a magnetic 'flea' for about 60min at room temperature. The precipitate which formed was collected on a Whatman No. 1 filter by filtration on a Buchner funnel and was then placed in a Buchner flask to which was added 160ml of 50mM-Tris/HCl (pH8.0)/0.02% $NaN_3$. Residual chloroform was removed under vacuum to prevent subsequent damage to centrifuge tubes. The mixture was then transferred to a Waring blender with 600ml of cold 5% (w/v) sodium deoxycholate in 50mM-Tris/HCl, pH8.0, and homogenized, followed by stirring overnight at 4°C. The soluble extract was then recovered by centrifugation at 35000rev./min for 30min in a Beckman type 35 rotor. The recovery of Thy-1 xenoantigenic activity, assayed as described in Barclay *et al.* (1975), from brain homogenate to deoxycholate extract was 48% with 4.4-fold purification.

The deoxycholate extract from two of the above preparations (thus from 400g wet wt. of rat brains; about 1.3 litre) was then passed through an M.R.C. OX 7 (anti-Thy-1.1) mouse monoclonal antibody affinity column of 20ml volume. The M.R.C. OX 7 IgG was purified as described in Mason & Williams (1980) and coupled to Sepharose 4B at 10mg/ml of beads which were activated with 20mg of CNBr/ml of beads by using the procedure of Porath (1974). The antibody was added to the activated beads at 13mg/ml in 50mM-sodium borate buffer, pH8.0, and almost all was covalently bound. Before use, the beads were washed with 50mM-diethylamine/HCl (pH11.5)/0.5% sodium deoxycholate, which is the buffer used for elution of bound antibody. In use, the antibody column was preceded by a 10ml column containing rabbit IgG coupled to Sepharose 4B to remove any aggregates in the extract. The deoxycholate extract was passed through the columns at 0.2ml/min and then the upper column was disconnected and the anti-Thy-1 column was washed until the $A_{280}$ was reduced to that of the buffer. The column was then eluted with 50mM-diethylamine (pH11.5)/0.5% sodium deoxycholate and the eluted fractions were immediately neutralized with solid glycine (Letarte-Muirhead *et al.*, 1975). The eluted material was concentrated and subjected to gel filtration on a Sephacryl S-200 column (3.2cm × 70cm) in 0.5% sodium deoxycholate.

In the affinity chromatography step the antigenic activity was removed from 1 litre of extract to 1% of the starting level and the column was saturated when 1.2 litres of extract had passed. In the elution step 60% of the activity removed from the extract was recovered, giving an overall yield from homogenate of 29%. The purification of antigenic activity per mg of protein was 770-fold compared with the starting brain homogenate. The eluted antigen contained minor contaminants, which were removed by gel filtration in which 95% of the antigenic activity loaded was recovered, and the purification factor was increased to 1000-fold. This is similar to the value of 720-fold obtained in the previous purification which used lentil-lectin affinity chromatography (Barclay *et al.*, 1975). No contaminants could be detected in the pure Thy-1 by SDS/polyacrylamide-gel electrophoresis and the pure material electrophoresed as a doublet on gels, as was noted previously (Kuchel *et al.*, 1978), when the acrylamide gel system of Laemmli (1970) was used. From the 400g of rat brain 10.7mg of pure Thy-1 was obtained, and this represents 22% of the theoretical capacity of an antibody column containing 200mg of anti-Thy-1 antibody. The column was re-used without detectable loss of activity. The Thy-1 was dialysed against 10mM-Tris/HCl (pH8.0)/0.02% $NaN_3$ to remove deoxycholate and then against distilled water before use in sequence studies.

With large amounts of pure Thy-1 an accurate determination of the u.v. absorption spectrum was possible. As can be seen in Fig. 1, the absorbance is

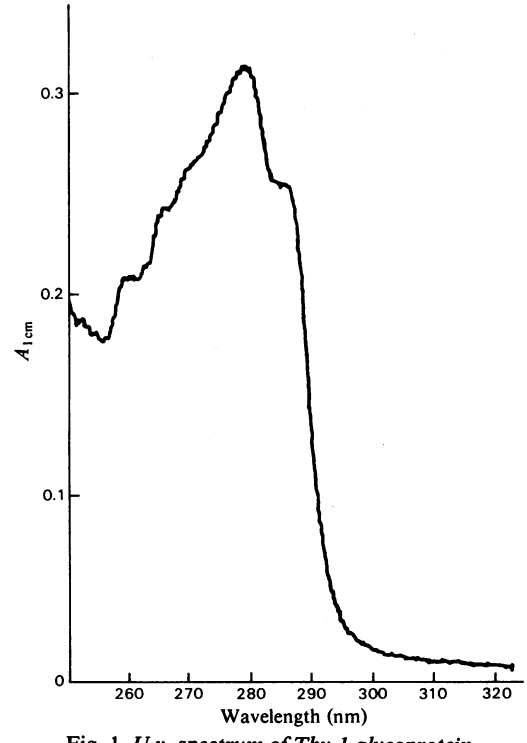


Fig. 1. *U.v. spectrum of Thy-1 glycoprotein*
The u.v. spectrum of Thy-1 glycoprotein at 1.2mg/ml in 0.5% deoxycholate/0.01M-Tris/HCl (pH8)/0.02% $NaN_3$ was read against the same buffer.

low and the $A_{1cm,278}$ value for a solution containing 10mg of glycoprotein/ml was $2.5 \pm 0.3$ (S.D.) from eleven determinations. The shape and magnitude of the spectrum shows that the molecule contains no tryptophan (Beavan & Holiday, 1952). The relationship between the weight of glycoprotein determined by amino acid analysis (normalized to include protein plus carbohydrate) and by the Lowry method was 1:1.02.

*Reduction and alkylation of rat brain Thy-1 glycoprotein*

Rat brain Thy-1 glycoprotein was dissolved in 7M-guanidinium chloride/0.5M-Tris/HCl, pH8.0, to give a concentration of 5–10mg/ml. Dithiothreitol was added to give a 40-fold molar excess over glycoprotein and the solution was stirred at 23°C for 1h. An 8-fold molar excess over glycoprotein of iodo[2-$^{14}$C]acetic acid (57mCi/mmol) was added, and after 2min unlabelled iodoacetic acid was added to give a 1.1-fold molar excess over all thiol groups. After mixing at 4°C for 30min the reaction was quenched with excess dithiothreitol. Excess reagents were then removed by dialysis against 0.1M-$NH_4HCO_3$.

*Succinylation of the reduced and alkylated Thy-1 glycoprotein*

Reduced and alkylated Thy-1 glycoprotein was dissolved in 1.0M-Tris/HCl buffer, pH10.5, to give a concentration of 10mg/ml. Amino groups were succinylated at 23°C by the addition of five portions of 20mg of succinic anhydride at 30min intervals. Excess reagents were removed by dialysis against 0.1M-$NH_4HCO_3$.

*Trypsin digestion of the succinylated, reduced and alkylated Thy-1 glycoprotein*

Succinylated, reduced and alkylated Thy-1 glycoprotein in 0.1M-$NH_4HCO_3$, pH8.0, at a concentration of 5mg/ml, was incubated for 6h at 37°C with 2% (w/w) Tos-Phe-$CH_2Cl$ treated trypsin. The digestion was ended by the addition of $iPr_2P$-F to a concentration of 6mM. The digest was then centrifuged at $1.2 \times 10^6$ *g*-min to remove any insoluble material prior to application to a column (1.4cm × 100cm) of Bio-Gel P-10 (200–400 mesh) equilibrated with 0.1M-$NH_4HCO_3$.

In some cases Brij 96 detergent was added at the outset to the trypsin digest at a concentration of 0.4% (v/v). The digested material was made 0.5M with respect to acetic acid and applied to a Bio-Gel P-10 column (1cm × 112cm) equilibrated with 0.1M-acetic acid.

*V-8 proteinase digestion of the reduced and alkylated Thy-1 glycoprotein*

Reduced and alkylated Thy-1 glycoprotein in 0.1M-$NH_4HCO_3$, pH7.8, at a concentration of 5mg/ml, was incubated for 20h at 37°C with 1.5% (w/w) V-8 proteinase which was added in two portions at 0h and 12h. The digestion was ended by heating at 100°C for 5min and then the digest was centrifuged at $1.2 \times 10^6$ *g*-min prior to application to a column (1.4cm × 100cm) of Bio-Gel P-10 (200–400 mesh) equilibrated with 0.1M-$NH_4HCO_3$.

*Pronase and pyroglutamate aminopeptidase digestion of peptides derived from the tryptic or V-8 proteinase digests*

*Pronase digestion.* Peptide T-8 (80nmol) in 0.25M-Tris/HCl/10mM-$CaCl_2$, pH7.4, was digested with Pronase for 72h at 37°C. The enzyme was added in three portions of 5% (w/w) at 0, 24 and 48h. The digestion was ended by heating at 100°C for 5min and the glycopeptide was then separated from other digestion products by gel filtration on a column (1.0cm × 100cm) of Bio-Gel P-6 (200–400 mesh) equilibrated with 0.1M-$NH_4HCO_3$.

*Pyroglutamate aminopeptidase digestion.* Peptide SP-1 in 50mM-sodium phosphate, pH7.4, at a concentration of 1mg/ml was treated with 40% (w/w) pyroglutamate aminopeptidase for 7h at 37°C and the digestion was ended by heating at 100°C for 5min. The peptide, with its blocked *N*-terminal amino acid now removed, was separated from the enzyme and salts by gel filtration on a column (1.0cm × 100cm) of Sephadex G-50 (fine grade) equilibrated with 50mM-$NH_4OH$, pH10.5. The pyroglutamate aminopeptidase was eluted near the void volume of the column well ahead of the unblocked peptide.

*Pepsin digestion of unreduced and reduced and alkylated Thy-1 glycoprotein*

Unreduced Thy-1 glycoprotein (160nmol) in 2.0ml of 0.5M-acetic acid was digested with pepsin (56$\mu$g in 200$\mu$l of 0.5M-acetic acid) at 37°C for 15h. The digest was applied to a column (1cm × 112cm) of Bio-Gel P-10 equilibrated with 0.1M-acetic acid and fractions (0.65ml) were collected.

Unreduced Thy-1 glycoprotein (95nmol) was also digested with pepsin [1% (w/w) added at 0, 6 and 22h during the digest], for 46h at 37°C, in the presence of 0.29% Brij 96 in 0.5M-acetic acid. The digest was fractionated in exactly the same manner as the previous digest.

Reduced and alkylated Thy-1 glycoprotein (17nmol) in 0.25ml of 0.5M-acetic acid was digested with pepsin (6$\mu$g in 10$\mu$l of 0.5M-acetic acid) at 37°C for 15h. The digest was fractionated in exactly the same manner as described for the digests of the unreduced Thy-1 glycoprotein samples.

*High-voltage paper electrophoresis*

High-voltage paper electrophoresis was per-

formed on No. 1 Whatman paper by the method of Katz *et al.* (1959) in the following buffers: pH 1.9, formic acid/acetic acid/water (2.5:7.8:89·7, by vol.); pH 3.5, pyridine/acetic acid/water (1:10:289, by vol.); pH 6.5, pyridine/acetic acid/water (5:112:1120, by vol.). After electrophoresis at 3 kV for approx. 1 h the peptides were stained with ninhydrin or by the starch–iodide method of Pan & Dutcher (1956). After preparative electrophoresis unstained peptides were eluted with 0.25 M-$NH_4OH$.

*Liquid-scintillation counting of radioactivity in column fractions*

Each sample (10–20 $\mu$l) of a column fraction was added to scintillation fluid (5 ml) composed of 2% (w/v) naphthalene/0.5% (w/v) 2,5-diphenyloxazole in 1,4-dioxan, or to Bio-Fluor (3 ml) (New England Nuclear).

*Manual dansyl–Edman degradation*

Manual dansyl–Edman degradation was performed by the methods of Hartley (1970) or Tarr (1978). The dansylated amino acids were identified by using t.l.c. as described by Woods & Wang (1967).

*Automated amino acid sequence determination*

Amino acid sequences were determined by automated Edman degradation in a Beckman 890C spinning-cup sequencer using a 0.1 M-Quadrol program (Brauer *et al.*, 1975) or a 0.3 M-Quadrol program (Hunkapiller & Hood, 1978). Polybrene (2 mg) was added to the sequencer cup prior to the application of each sample (Klapper *et al.*, 1978). The thiazolinones released were converted into the amino acid phenylthiohydantoin derivatives by heating at 80°C under $N_2$ for 10 min in 1 M-HCl (200 $\mu$l) containing 0.1% ethanethiol. After two extractions with ethyl acetate (700 $\mu$l) the extracted amino acid phenylthiohydantoin derivatives were identified by t.l.c. (Summers *et al.*, 1973), back hydrolysis and amino acid analysis (Mendez & Lai, 1975) or by high-pressure liquid chromatography. The latter was performed on a $\mu$-Bondapak C-18 column with a 14–56% methanol gradient in aqueous buffer (Bridgen *et al.*, 1976) pumped at 2.2 ml/min by a Waters Associates system (model 6000A) with a u.v. detector. An isocratic 27% methanol elution was used to identify the basic amino acid phenylthiohydantoin derivatives left in the acid phase. *S*-[$^{14}$C]Carboxymethylcysteine was also identified by counting radioactivity in the butyl chloride extracts.

*Amino acid and hexosamine analysis*

*Amino acid analysis.* Reduced and alkylated Thy-1 glycoprotein was hydrolysed, under vacuum, at 110°C for 24, 48 and 72 h in twice glass-distilled 5.7 M-HCl which contained 5 mM-phenol and 0.05% 2-mercaptoethanol. The serine and threonine values were extrapolated back to zero time of hydrolysis, the valine and isoleucine values were taken as found in the 72 h hydrolysate and the values for the other amino acids were averaged over the three hydrolysis times. Peptides were hydrolysed under $N_2$ for 24 h at 110°C in twice glass-distilled 5.7 M-HCl, 5 mM in phenol; no correction was made for destruction of serine or threonine.

*Hexosamine analysis.* Hexosamine analysis was performed by the method of Allen & Neuberger (1975) in which the sample is heated in 3 M-*p*-toluenesulphonic acid at 110°C for 24 h. Recoveries of glucosamine and galactosamine were 87% and 70% respectively.

All samples were analysed on a Durrum D-500 amino acid analyser.

## Results

The amino acid sequence of rat brain Thy-1 glycoprotein was obtained from the studies performed on the peptides generated by digestion of the entire glycoprotein with either trypsin or V-8 proteinase. The peptides were purified by an initial gel-filtration step, followed by paper electrophoresis when necessary.

*Isolation and amino acid sequencing of the peptides derived from the tryptic digest of Thy-1 glycoprotein*

The digest of the reduced and alkylated, succinylated Thy-1 glycoprotein was centrifuged at $1.2 \times 10^6$ *g*-min to remove insoluble material. Amino acid analysis showed that the pellet contained less than 0.4% of the initial protein and that it did not have a hydrophobic composition as would be expected if the tryptic cleavage of Thy-1 glycoprotein had left an insoluble 'core'. The supernatant was fractionated by gel filtration on a column (1.4 cm × 100 cm) of Bio-Gel P-10 (200–400 mesh) equilibrated with 0.1 M-$NH_4HCO_3$, as shown in Fig. 2.

*Peptide T-1* [residues 1–2 (full sequence numbering; Fig. 3)]. No free amino-terminal amino acid could be detected for peptide T-1 by the dansyl method. From its amino acid composition (Table 1), and the known specificity of trypsin, it appeared that the amino acid sequence of peptide T-1 was (Glx)-Arg. Peptide T-1 was neutral on paper electrophoresis at pH 6.5 and this, together with the absence of a free amino-terminus, suggested that the Glx was pyroglutamic acid. The presence of an *N*-terminal pyroglutamic acid is consistent with the finding that intact Thy-1 glycoprotein gave no sequence when analysed with the automatic sequencer (Kuchel *et al.*, 1978) and this was further confirmed by studies on SP-1 (below).

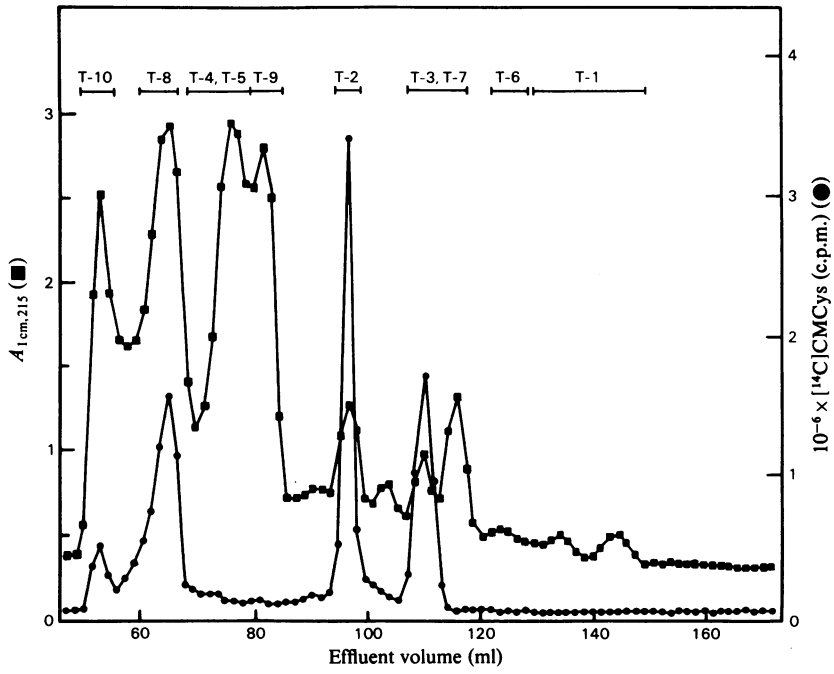


Fig. 2. *Fractionation of the tryptic digest of Thy-1 glycoprotein on Bio-Gel P-10*

The supernatant (2.0ml) from the trypsin digest of rat brain Thy-1 glycoprotein (10mg) was applied to a column (1.4cm × 100cm) of Bio-Gel P-10, which was equilibrated with 0.1M-$NH_4HCO_3$ and run at room temperature. The column was run at 5ml/h and 1.5ml fractions were collected. Subsequently the fractions within the horizontal bars were pooled and freeze dried. ■, $A_{1cm,215}$; ●, radioactivity (c.p.m.).

*Peptide T-2* (residues 3–16). The entire amino acid sequence of this peptide was obtained by automated Edman degradation (Fig. 3; Table 1).

*Peptide T-3* (residues 17–20). The sequence of this peptide was determined by manual dansyl–Edman degradation (Fig. 3; Table 1).

*Peptide T-4* (residues 21–37). Automated Edman degradation of peptide T-4 enabled identification of all residues except those at positions 23, 34 and 37 (full sequence numbering) (Fig. 3; Table 1). The presence of the carboxy-terminal arginine was inferred from the known specificity of trypsin and the serine was placed at 34 on the basis of the composition of a V-8 proteinase peptide derived from T-4 (data not shown). The other unidentified residue at position 23 was concluded to be Asx by comparison of the amino acid composition of peptide T-4 with the residues identified by sequence analysis (Table 1). As peptide T-4 contained 1.6mol/mol of glucosamine (Table 2), the presence of an asparagine-linked carbohydrate group is indicated for position 23. This conclusion is supported by the fact that the sequence Asn-Asn-Thr would be compatible with attachment of glucosamine to Asn-23. The yield of phenylthiohydantoin amino acids from the automated Edman degradation did not drop after this residue, which showed that the reaction and cleavage of the residue, during the sequencer cycle, had occurred normally.

*Peptide T-5* (residues 38–56). The entire amino acid sequence of this peptide was obtained by automated Edman degradation (Fig. 3; Table 1).

*Peptide T-6* (residues 57–58). One cycle of manual dansyl–Edman degradation identified the *N*-terminal serine and released free arginine.

*Peptide T-7* (residues 59–65). The first six residues were identified by manual dansyl–Edman degradation and the carboxy-terminal position of the arginine was deduced from the known specificity of trypsin. Thin-layer chromatography of the phenylthiohydantoin derivative released at position 2 in peptide T-7 established that this residue was asparagine (residue 60). A net charge of close to zero on electrophoresis of peptide T-7 at pH6.5 established that aspartic acid was present at position 6 (residue 64).

*Peptide T-8* (residues 66–88). Automated Edman degradation of peptide T-8 enabled identification of 22 of the 23 residues (Fig. 3; Table 1). This sequence was also consistent with the analysis and *N*-termini

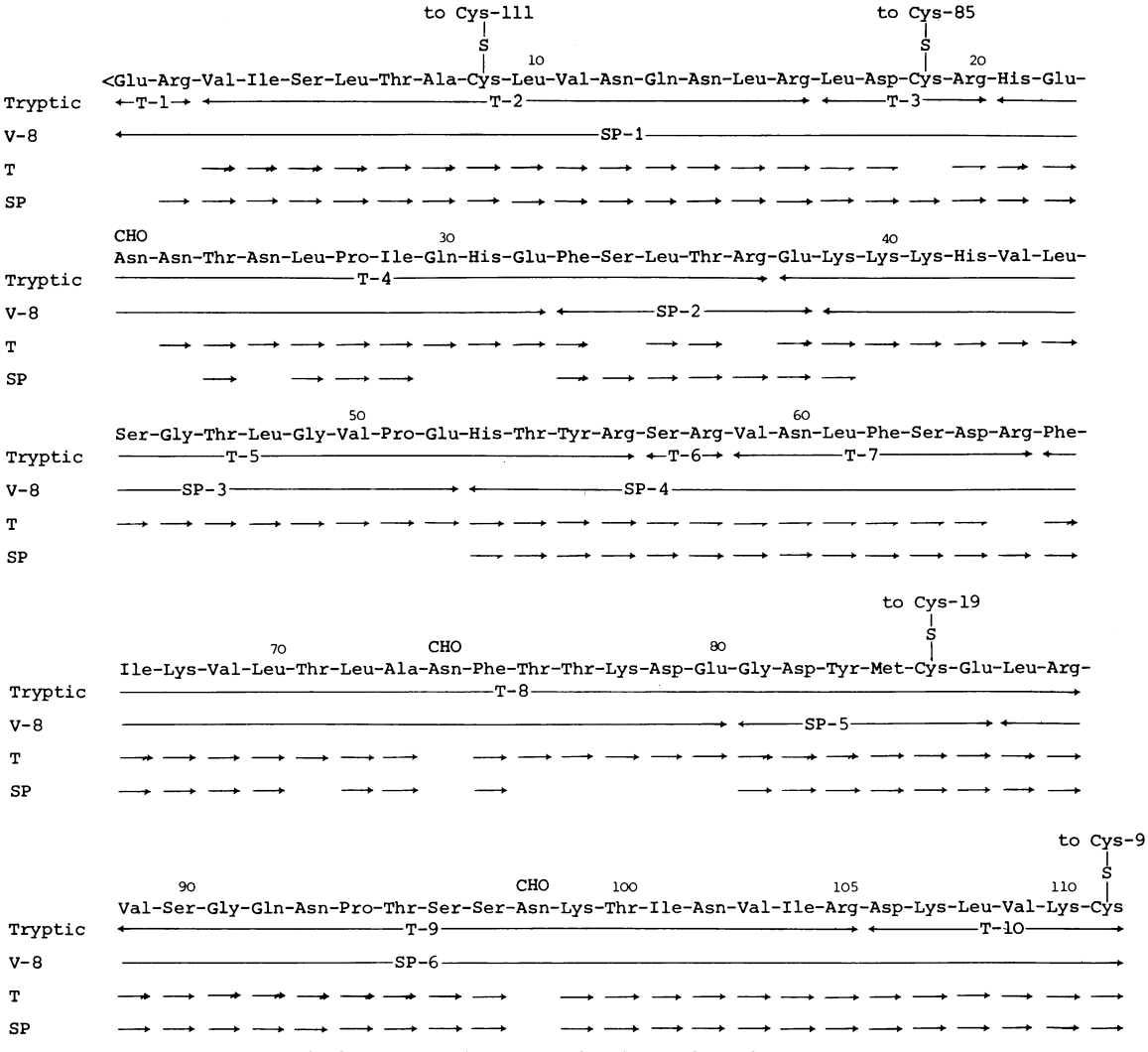


Fig. 3. *Amino acid sequence of rat brain Thy-1 glycoprotein*

Peptides prefixed by T- and SP- were derived from tryptic and V-8 proteinase digests respectively. The *N*-terminal amino acid of the intact molecule was judged to be 5-oxopyrrolidine-2-carboxylic acid for the reasons outlined in the text. Sequence determination from the *N*-terminus by use of the Beckman 890C protein sequencer, or by the dansyl procedure, or by both techniques, is shown by →, ⇀ and ⇻ respectively. Positions 22, 74 and 98 are probably occupied by asparagine residues that have carbohydrate attached. The reasons for these assignments are given in the text. Abbreviation used: CHO, carbohydrate.

of peptides prepared from T-8 by V-8 proteinase digestion and CNBr cleavage [data not shown; see Campbell (1978)]. No residue was identified at position 9 (residue 74 in the full sequence) but comparison of the amino acid composition with the amino acids identified during the automated Edman degradation indicated that one Asx residue was unaccounted for. As peptide T-8 contained 2.6 residues of glucosamine (Table 2), the presence of an asparagine-linked carbohydrate group is indicated. The sequence Asn-Phe-Thr is compatible with glycosylation of Asn-74. Digestion of peptide T-8 with Pronase and separation of the glycopeptide from small peptides by gel filtration produced a glycopeptide whose main amino acid constituent was aspartic acid (composition was: Asx, 1.0; Ala, 0.6; Lys, 0.5; Glx, 0.4; Gly, 0.4; all others $\leqslant$0.1). This provided further evidence that residue 74 contained an asparagine-linked carbohydrate chain.

*Peptide T-9* (residues 89–105). Automated Ed-

Table 1. *Amino acid compositions and details of the automated Edman degradation of the tryptic peptides derived from Thy-1 glycoprotein*

Compositions are given as mol of residue/mol of peptide and values less than 0.2 are omitted. All values shown have been normalized to Arg = 1.0. Abbreviation: N.Q., not quantified. The values given in parentheses are those obtained by amino acid sequence analysis. No corrections were made for destruction of serine and threonine.

| | Amino acid composition (mol of residue/mol) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Peptide ... | T-1 | T-2 | T-3 | T-4 | T-5 | T-6 | T-7 | T-8 | T-9 | T-10 |
| Asx | — | 1.9 (2) | 1.2 (1) | 2.9 (3) | 0.3 (0) | — | 1.9 (2) | 3.1 (3) | 3.0 (3) | 1.0 (1) |
| Thr | — | 0.9 (1) | — | 1.8 (2) | 1.9 (2) | — | — | 2.5 (3) | 1.8 (2) | 0.3 (0) |
| Ser | 0.3 | 1.2 (1) | 0.4 (0) | 1.0 (1) | 1.0 (1) | 1.4 (1) | 1.0 (1) | 1.0 (0) | 2.6 (3) | 0.5 (0) |
| Glx | 1.1 (1) | 1.4 (1) | — | 2.9 (3) | 2.1 (2) | 0.9 (0) | — | 2.4 (2) | 1.4 (1) | 0.4 (0) |
| Pro | — | — | — | 1.0 (1) | 1.0 (1) | — | — | — | 1.0 (1) | — |
| Gly | — | 0.7 (0) | 0.3 (0) | 0.5 (0) | 2.2 (2) | — | 0.3 (0) | 1.5 (1) | 1.8 (1) | 0.4 (0) |
| Ala | 0.4 | 1.0 (1) | — | — | — | 0.3 (0) | — | 1.2 (1) | 0.4 (0) | 0.2 (0) |
| Cys | N.Q. (0) | N.Q.* (1) | N.Q.* (1) | N.Q. (0) | N.Q. (0) | N.Q. (0) | N.Q. (0) | N.Q.* (1) | N.Q. (0) | 0.7* (1) |
| Val | — | 1.4 (2) | 0.2 (0) | — | 2.1 (2) | 0.2 (0) | 0.9 (1) | 1.4 (1) | 1.6 (2) | 1.0 (1) |
| Met | — | — | — | — | — | — | — | 0.2 (1) | — | — |
| Ile | — | 0.6 (1) | — | 0.8 (1) | — | — | — | 1.4 (1) | 1.3 (2) | 0.2 (0) |
| Leu | — | 2.5 (3) | 1.1 (1) | 1.8 (2) | 1.9 (2) | — | 0.9 (1) | 2.8 (3) | 0.3 (0) | 0.9 (1) |
| Tyr | — | — | — | — | 1.0 (1) | — | — | 0.7 (1) | — | 0.2 (0) |
| Phe | — | — | — | 1.1 (1) | — | — | 0.9 (1) | 1.5 (2) | — | — |
| His | — | — | — | 1.7 (2) | 1.9 (2) | — | — | — | — | 0.1 (0) |
| Lys | — | 0.5 (0) | — | — | 2.8 (3) | — | — | 1.9 (2) | 1.0 (1) | 1.4 (2) |
| Arg | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 1.0 (1) | 0.1 (0) |
| Total residues | 2 | 14 | 4 | 17 | 19 | 2 | 7 | 23 | 17 | 6 |
| Yield of peptide (%) | 49 | 100 | 54 | 30 | 23 | 50 | 55 | 100 | 68 | 70 |
| Amount (nmol) used in automated sequence analysis | — | 180 | — | 85 | 180 | — | — | 200 | 175 | 55 |
| Initial yield (%) | — | 13 | — | 6 | 24 | — | — | 8 | 6 | 53 |
| Repetitive yield (%) | — | 94 | — | 88 | 94 | — | — | 94 | 90 | 62 |

* Indicates the presence of *S*-[$^{14}$C]carboxymethylcysteine determined by radioactivity.

Table 2. *Amino sugar composition of tryptic and V-8 proteinase peptides derived from rat brain Thy-1 glycoprotein*

| | Amount of amino sugar (mol/mol of peptide) | |
|---|---|---|
| Peptide | Glucosamine | Galactosamine |
| T-4 | 1.6 | 0 |
| SP-1 | 2.3 | 0 |
| T-8 | 2.6 | 0 |
| SP-4 | 3.0 | 0 |
| T-9 | 2.6 | 0 |
| SP-6 | 2.6 | 0.7 |
| T-10 | 0.7 | 0.7 |
| Brain Thy-1* | 9.2 | 1.1 |

* From Barclay *et al.* (1976).

man degradation allowed identification of 14 of the first 15 residues of peptide T-9 (Fig. 3; Table 1). Residue 10 (98 in the full sequence) was not identified but this is likely to be Asn with *N*-linked glucosamine attached. This conclusion was drawn since one Asx was unaccounted for in the composition. The peptide contained 2.6 mol of glucosamine/mol (Table 2) and the sequence Asn-Lys-Thr would be compatible with a signal sequence for attachment of glucosamine.

*Peptide T-10* (residues 106–111). A satisfactory yield of this peptide was not achieved when the digest was fractionated as in Fig. 2, although an *N*-terminal aspartic residue was determined by the dansyl technique from the front fractions, which are designated T-10. The problem of obtaining T-10 was overcome by including the detergent Brij 96 in the tryptic digest, with subsequent fractionation on Bio-Gel P-10 in 0.1 M-acetic acid. Using this fractionation without Brij 96, only three $^{14}$C-labelled tryptic peptides were obtained and no peptides were detected at the column front (Fig. 4). The inclusion of Brij 96 gave a peak of $^{14}$C at the column front, with a yield of about 50% compared with the other three peptides (Fig. 4). The front fraction had the composition shown in Table 1 and T-10 must be the

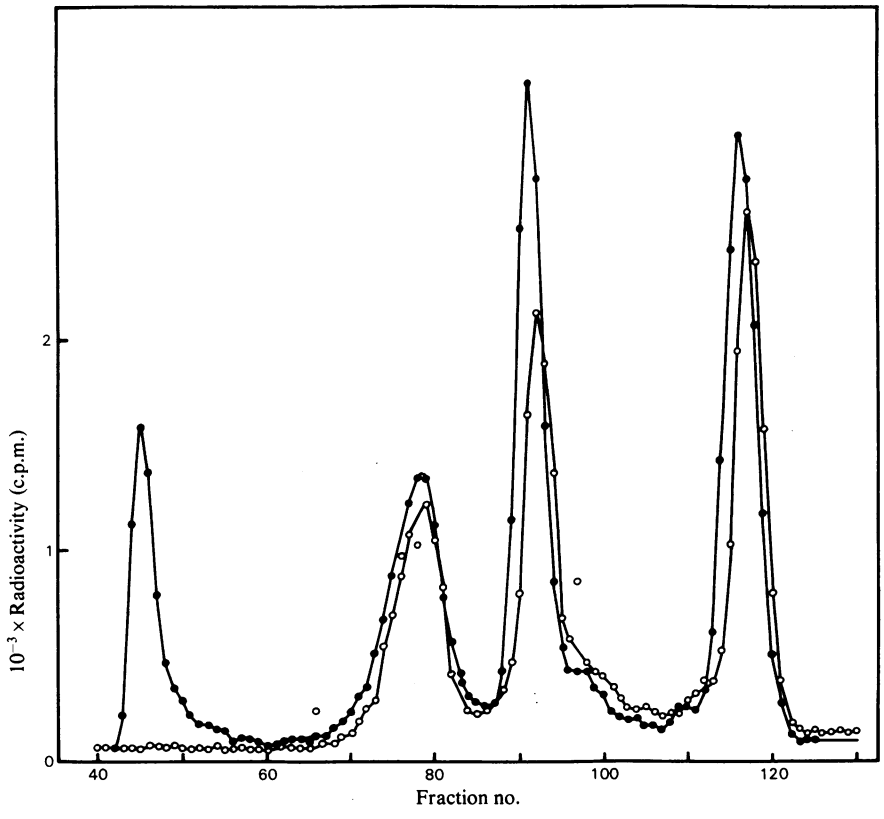


Fig. 4. *Effect of Brij 96 detergent on gel filtration of tryptic peptides*
*S*-[$^{14}$C]Carboxymethylated and succinylated Thy-1 glycoprotein (10nmol) was digested at 37°C for 22h with trypsin in the absence (○) or presence (●) of 0.4% (w/w) Brij 96 detergent. Peptides were separated on a Bio-Gel P-10 column in 0.1M-acetic acid. Fractions (0.65ml) were collected and $^{14}$C-radioactivity (c.p.m.) was counted on a 50$\mu$l sample.

*C*-terminal peptide, since it did not contain arginine residues. In addition to the residues shown in Table 1, amino acid analysis of T-10 showed an unidentified peak between lysine and $NH_3$. Quantitation of this, using the colour yield for lysine, gave 1.2 residues/mol of peptide. The elution position of this unidentified residue was the same as that of ethanolamine.

Automated Edman degradation gave the sequence shown in Fig. 3. Both lysine residues in T-10 were detected as phenylthiohydantoin succinyl-lysine and thus the unidentified material cannot be attached to the ε-amino group of either lysine. Cys-111 was detected in the sequence by $^{14}$C radioactivity and as phenylthiohydantoin *S*-carboxymethylcysteine at a yield compared with the previous residue of 40%. No residue was detected after Cys-111 and the sequence and composition are consistent with this being the last conventional amino acid. However, if the unidentified ninhydrin-positive material is an unusual amino acid, then this, rather than Cys-111, would be the *C*-terminal residue.

Analysis for hexosamines showed that T-10 contained about 1mol each of glucosamine and galactosamine per mol of peptide (Table 2).

*Isolation and amino acid sequencing of the peptides derived from the digestion of Thy-1 glycoprotein with V-8 proteinase*

The digest of the reduced and alkylated Thy-1 glycoprotein was centrifuged at $1.2 \times 10^6$ *g*-min and the supernatant was fractionated on a column (1.4cm × 100cm) of Bio-Gel P-10 (200–400 mesh) equilibrated with 0.1M-$NH_4HCO_3$ (Fig. 5). Further purification of the fractions pooled as indicated in Fig. 5 was performed, if necessary, by paper electrophoresis. The amino acid and hexosamine compositions of the peptides produced by digestion with V-8 proteinase are shown in Tables 3 and 2 respectively.

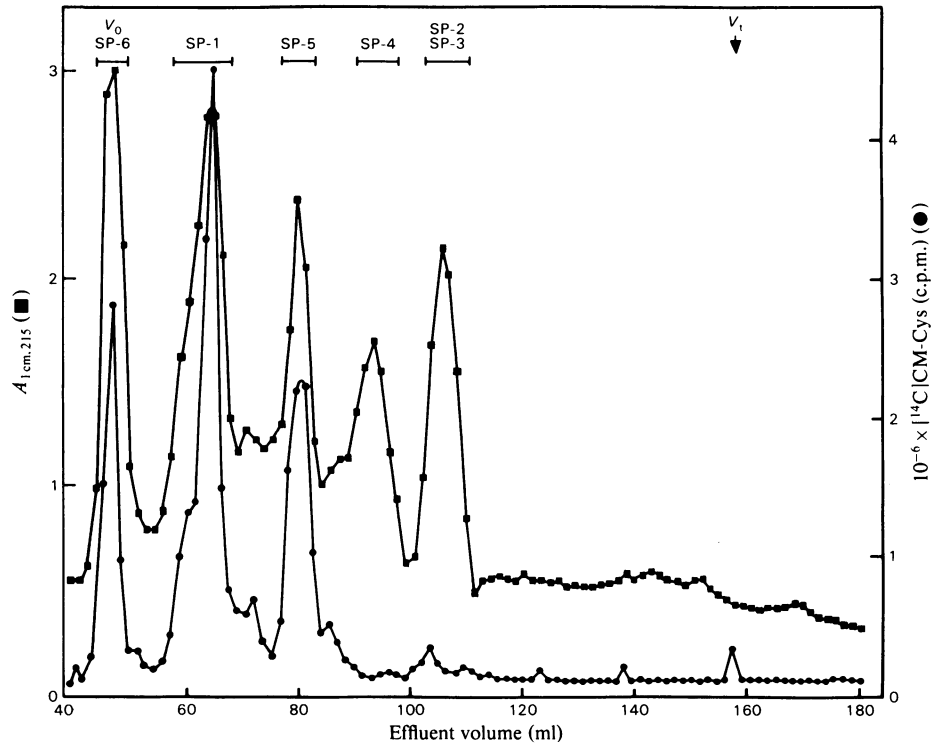


Fig. 5. *Fractionation of the V-8 proteinase digest of Thy-1 glycoprotein on Bio-Gel P-10*

The supernatant (2.0ml) from the V-8 proteinase digest of rat brain Thy-1 glycoprotein (10mg) was applied to a column (1.4cm × 100cm) of Bio-Gel P-10, which was equilibrated with 0.1M-$NH_4HCO_3$ and run at room temperature. The column was run at 5ml/h and 1.5ml fractions were collected. Subsequently the fractions within the horizontal bars were pooled and freeze-dried. ■, $A_{1cm,215}$; ●, radioactivity (c.p.m.).

*Peptide SP-1* (residues 1–32). No free amino-terminus group could be detected for peptide SP-1 by the dansyl method; therefore peptide SP-1 was treated with pyroglutamate aminopeptidase as described in the Experimental section and then subjected to automated Edman degradation. Sequence data were obtained which showed that the enzyme had 'unblocked' the peptide and thus confirmed the presence of a pyroglutamyl residue at the *N*-terminus of peptide SP-1. Automated Edman degradation allowed identification of the first 21 amino acid residues and four of the remaining ten residues (Fig. 3). These sequence data, when taken together with the known tryptic peptide sequences, allowed the alignment of the tryptic peptides T-1, T-2, T-3, and T-4 (Fig. 3). Peptide SP-1 contained 2.3mol/mol of glucosamine (Table 2) and from the studies on peptide T-3 it was known that this carbohydrate was present at residue 23. The presence of a glutamic acid residue at position 22, which had not been cleaved by the V-8 proteinase, suggests that the carbohydrate moiety at position 23 caused steric hindrance, thus preventing cleavage.

*Peptide SP-2* (residues 33–38). Automated Edman degradation gave the full sequence of this peptide (Fig. 3; Table 1). The determination of Glu-38 provided the overlap between peptides T-4 and T-5.

*Peptide SP-3* (residues 39–52). This peptide was not sequenced other than the determination of an *N*-terminal lysine by the dansyl method. The composition (Table 3) was consistent with the sequence of residues 39–52 within T-5.

*Peptide SP-4* (residues 53–80). Automated Edman degradation allowed identification of 21 of the 28 amino acid residues present in peptide SP-4 (Fig. 3; Table 3). The presence of glutamic acid at position 28 (full sequence 80) was concluded from the known specificity of the V-8 proteinase and the amino acid sequence of peptide T-8. The residue at position 22 (residue 74) was the asparagine-linked carbohydrate group which was identified at this position in peptide T-8. The partial amino acid sequence obtained for peptide SP-4 allowed the alignment of the tryptic peptides T-5, T-6, T-7 and T-8.

Table 3. *Amino acid compositions and details of the automated Edman degradation of the V-8 proteinase peptides derived from Thy-1 glycoprotein*

Compositions are given as mol of residue/mol of peptide and values less than 0.2 are omitted. Values shown for any particular peptide are normalized to the residue marked with †. Abbreviation: N.Q., not quantified. The values given in parentheses are those obtained by amino acid sequence analysis. No corrections were made for destruction of serine and threonine.

| | Amino acid composition (mol of residue/mol) | | | | | |
|---|---|---|---|---|---|---|
| Peptide ... | SP-1 | SP-2 | SP-3 | SP-4 | SP-5 | SP-6 |
| Asx | 6.1 (6) | 0.3 (0) | — | 3.9 (4) | 1.2 (1) | 3.8 (4) |
| Thr | 2.1 (2) | 0.9 (1) | 1.1 (1) | 3.2 (4) | 0.3 (0) | 2.0 (2) |
| Ser | 1.0 (1) | 0.9 (1) | 1.0 (1) | 2.2 (2) | 0.2 (0) | 2.7 (3) |
| Glx | 4.7 (5) | 1.1 (1) | 1.2 (1) | 1.6 (1) | 1.0† (1) | 1.6 (1) |
| Pro | 1.0 (1) | — | 1.1 (1) | — | — | 1.2 (1) |
| Gly | 0.4 (0) | — | 2.3 (2) | 0.9 (0) | 0.8 (1) | 1.8 (1) |
| Ala | 1.0 (1) | — | — | 1.0 (1) | 0.2 (0) | 1.0 (0) |
| Cys | N.Q.* (2) | N.Q. (0) | N.Q. (0) | N.Q. (0) | N.Q.* (1) | 0.7* (1) |
| Val | 1.5 (2) | — | 2.1 (2) | 2.0 (2) | 0.2 (0) | 2.9 (3) |
| Met | — | — | — | — | 0.1 (1) | — |
| Ile | 1.5 (2) | — | — | 1.0 (1) | — | 1.7 (2) |
| Leu | 4.5 (5) | 1.0 (1) | 2.2 (2) | 2.5 (3) | 0.3 (0) | 2.3 (2) |
| Tyr | — | — | — | 0.4 (1) | 0.6 (1) | 0.3 (0) |
| Phe | — | 0.9 (1) | — | 1.7 (3) | — | 0.4 (0) |
| His | 1.6 (2) | — | 1.0 (1) | 0.8 (1) | 0.2 (0) | 0.4 (0) |
| Lys | 0.4 (0) | — | 3.0† (3) | 1.7 (2) | 0.2 (0) | 3.0† (3) |
| Arg | 3.0† (3) | 1.0† (1) | — | 3.0† (3) | 0.3 (0) | 1.7 (2) |
| Total residues | 32 | 6 | 14 | 28 | 6 | 25 |
| Yield of peptide (%) | 58 | 45 | 33 | 31 | 112 | 70 |
| Amount (nmol) used in automated sequence analysis | 150 | 60 | — | 100 | 80 | 150 |
| Initial yield (%) | 3.3 | 50 | — | 5.5 | 38 | 73 |
| Repetitive yield (%) | 89 | 82 | — | 90 | — | 90 |

* Indicates the presence of *S*-[$^{14}$C]carboxymethylcysteine determined by radioactivity.

*Peptide SP-5* (residues 81–86). All six amino acid residues of this peptide were identified by automated Edman degradation (Fig. 3; Table 3).

*Peptide SP-6* (residues 87–111 plus *C*-terminal portion). This peptide was unusual in its elution volume (Fig. 5) and appeared to be aggregated, presumably due to the hydrophobic nature of the *C*-terminus (see peptide T-10 and pepsin peptides). Automated Edman degradation allowed the identification of 24 of the first 25 amino acid residues present in peptide SP-6 (Fig. 3). The amino acid sequence obtained for the *N*-terminal region of peptide SP-6 showed that peptide T-9 followed peptide T-8. Position 12 of peptide SP-6 (full sequence residue 98) was not identified and this corresponds to position 10 of peptide T-9 which is likely to be asparagine with an *N*-linked carbohydrate group attached (see above). One Asx residue in the SP-6 composition was unaccounted for in the sequence and the peptide contains glucosamine (Table 2).

The amino acid composition of peptide SP-6 (Table 3) indicates the presence of some alanine, glycine and Glx which is not accounted for. The glycine and Glx values are probably not significant, since high values for these residues were obtained in analyses of some other V-8 proteinase peptides. The alanine analysis suggests there may be one residue of this amino acid beyond the *S*-carboxymethylcysteine residue located at position 111 (Fig. 3) but this is not supported by the analysis of the *C*-terminal tryptic peptide T-10 (Table 1).

Acid hydrolysates of peptide SP-6, in addition to the amino acid composition shown in Table 3, contained an unidentified ninhydrin-positive peak which eluted between lysine and $NH_3$ on amino acid analysis. If the colour value for lysine is used to estimate the amount of this material, then it can be concluded that there is 1.5 mol of the unidentified ninhydrin positive material/mol of peptide SP-6. Furthermore, peptide SP-6 was found to contain galactosamine at 0.7 residues/mol.

### *Determination of the positions of the disulphide bonds in the Thy-1 glycoprotein*

Amino acid analysis of the reduced and alkylated Thy-1 glycoprotein showed that there are four *S*-carboxymethylcysteine residues per molecule

Table 4. *Amino acid composition of rat brain Thy-1 glycoprotein as determined by total amino acid analysis and by amino acid sequence analysis*

Samples of reduced and alkylated rat brain Thy-1 glycoprotein were hydrolysed at 110°C for 24, 48 and 72h. Experimental details are given in the text. Results are expressed as mol of amino acid/mol of protein, taking a value of 12500 for the molecular weight of the polypeptide portion of the Thy-1 glycoprotein.

| | Residues/mol | |
|---|---|---|
| | Total amino acid analysis | Taken from amino acid sequence of residues 1–111 |
| Asx | 14.9 | 15 |
| Thr | 9.3 | 10 |
| Ser | 7.3 | 8 |
| Glx | 10.8 | 10 |
| Pro | 3.1 | 3 |
| Gly | 4.9 | 4 |
| Ala | 2.8 | 2 |
| Cys | 4.3 | 4 |
| Val | 9.2 | 9 |
| Met | 1.0 | 1 |
| Ile | 4.6 | 5 |
| Leu | 12.7 | 13 |
| Tyr | 2.3 | 2 |
| Phe | 3.8 | 4 |
| His | 4.1 | 4 |
| Lys | 8.1 | 8 |
| Arg | 8.5 | 9 |

(Table 4). No evidence has been obtained for the presence of free sulphydryl groups {e.g. by incubation of unreduced Thy-1 glycoprotein with [$^{14}$C]iodoacetic acid in the presence of 7M-guanidinium chloride (Campbell, 1978)}; therefore it can be concluded that there are probably two disulphide bridges present per molecule of Thy-1 glycoprotein. Examination of pepsin digests of reduced and alkylated, and unreduced, Thy-1 glycoprotein has allowed the positions of these two disulphide bonds to be established.

A pepsin digest of *S*-[$^{14}$C]carboxymethylated Thy-1 glycoprotein was examined so that the elution positions of the peptic peptides containing the *S*-carboxymethylcysteine residues located at residues 9, 19, 85 and 111 could be determined. Fractionation of a pepsin digest of reduced and alkylated Thy-1 glycoprotein on a column (1cm × 112cm) of Bio-Gel P-10, in 0.1M-acetic acid, yielded three major peaks of radioactivity at 54, 70 and 78ml. The radioactively labelled peptides were further purified by electrophoresis at pH6.5 and located by autoradiography. Three major radiolabelled peptides were found with amino acid compositions: (1) CMCys, 1.0; Asx, 3.2; Thr, 1; Glx, 1.1; Leu, 0.7; His, 0.6; Arg, 0.4; (2) CMCys, 1.0; Thr, 0.8; Ala, 1.0; Leu, 1.0; (3) CMCys, 0.8; Glx, 1.0. It can be seen that these peptic peptides are derived from the amino acid sequences 17–25, 7–10 and 85–86 around the *S*-carboxymethylcysteine residues located at positions 19, 9 and 85 respectively in the intact Thy-1 glycoprotein (Fig. 3). No peptide could be isolated from this pepsin digest which corresponded to the amino acid sequence around the *S*-carboxymethylcysteine residue located at position 111 (Fig. 3). The loss of this peptide is analogous to the loss of the *C*-terminal tryptic peptide unless Brij 96 was added in the fractionation (Fig. 4).

Unreduced Thy-1 glycoprotein was digested with pepsin and the digest fractionated on a column (1cm × 112cm) of Bio-Gel P-10, in 0.1M-acetic acid, in exactly the same manner as for the pepsin digest of the reduced and alkylated material. No significant amount of peptide material was eluted until 39ml (approx. 9ml after the void volume of the column). All the peptide fractions eluted from the column were oxidised with performic acid and amino acid analysis showed the presence of cysteic acid only in the material eluted from 48–55ml. No smaller peptide fractions containing cysteic acid were found before, or after, oxidation with performic acid. The peptide fractions eluted in the 48–55ml region of the Bio-Gel P-10 column were pooled and further purified by paper electrophoresis at pH1.9. One major peptide fraction was obtained which was found to have an amino acid composition (Table 5) consistent with it being composed of approximately equimolar amounts of peptides derived from residues 16–27 and residues 83–86, i.e. the amino acid sequences around the half-cystine residues located at 19 and 85 respectively (Fig. 3). When this peptide fraction was oxidised with performic acid and reapplied to a column (1cm × 112cm) of Bio-Gel P-10, in 0.1M-acetic acid, three major peptide fractions were obtained, at 53, 83 and 94ml, which had amino acid compositions consistent with their being derived from residues 16–27, 84–86 and 83–86 respectively of the intact Thy-1 glycoprotein (Table 5; Fig. 3). These results establish that the half-cystine residue at position 19 is linked to the half-cystine residue at position 85.

The peptides containing the half-cystine residues at positions 9 and 111 could not be identified amongst the peptides obtained from the pepsin digest of unreduced Thy-1. The peptic peptide containing Cys-111 was also lost when the reduced and alkylated digest was fractionated (see above) and it thus seemed possible that in the unreduced digest peptides containing Cys-9 and Cys-111 were lost together. In an attempt to obtain these peptides in good yield, Brij 96 detergent was added during the pepsin digestion of unreduced Thy-1 glycoprotein.

Table 5. *Amino acid compositions of the disulphide-linked peptides isolated from the pepsin digest of non-reduced Thy-1 glycoprotein*

The values given in parentheses are those expected from the known amino acid sequence. Half-cystine and methionine were determined as cysteic acid and methionine sulphone respectively. Abbreviation used: N.D., not determined due to the presence of sugar residues. — denotes less than 0.2 mol of residue/mol, (–) denotes no residue.

| | Amino acid composition (mol of residues/mol) | | | | | | |
|---|---|---|---|---|---|---|---|
| Peptide ... | Cys-19–Cys-85 | | | | Cys-9–Cys-111 | | |
| | Before oxidation | After oxidation | | | Before oxidation | After oxidation | |
| Elution vol. (ml) | 48–55 | 53 | 83 | 94 | Void | Void | 70.5 |
| Asx | 3.8 (4) | 4.4 (4) | 0.2 (–) | 0.2 (–) | 0.2 (–) | – (–) | – (–) |
| Thr | 0.9 (1) | 1.1 (1) | – (–) | – (–) | 1.0 (1) | 0.2 (–) | 1.0 (1) |
| Ser | 0.3 (–) | – (–) | 0.2 (–) | 0.2 (–) | – (–) | – (–) | – (–) |
| Glu | 2.2 (2) | 1.3 (1) | 1.3 (1) | 1.3 (1) | – (–) | – (–) | – (–) |
| Pro | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) |
| Gly | 0.5 (–) | – (–) | 0.4 (–) | 0.4 (–) | – (–) | – (–) | – (–) |
| Ala | – (–) | – (–) | – (–) | – (–) | 1.3 (1) | 0.2 (–) | 1.2 (1) |
| Cys | 1.5 (2) | 0.9 (1) | 0.9 (1) | 1.0 (1) | 1.8 (2) | 1.0 (1) | 1.0 (1) |
| Val | 0.2 (–) | – (–) | – (–) | 0.2 (–) | 1.0 (1) | 0.9 (1) | – (–) |
| Met | 0.8 (1) | – (–) | 0.9 (1) | 1.2 (1) | – (–) | – (–) | – (–) |
| Ile | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) |
| Leu | 1.8 (2) | 2.1 (2) | – (–) | – (–) | 1.1 (1) | 0.2 (–) | 1.0 (1) |
| Tyr | N.D. (–) | – (–) | – (–) | 1.1 (1) | – (–) | – (–) | – (–) |
| Phe | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) | – (–) |
| His | 0.9 (1) | 1.1 (1) | – (–) | – (–) | – (–) | – (–) | – (–) |
| Lys | – (–) | – (–) | – (–) | – (–) | 1.0 (1) | 1.1 (1) | – (–) |
| Arg | 1.5 (2) | 1.9 (2) | – (–) | – (–) | – (–) | – (–) | – (–) |
| Yield (%) | 40 | 28 | 28 | 28 | 69 | 54 | 43 |
| Expected position in amino acid sequence (residues) | 16–27 +83–86 | 16–27 | 84–86 | 83–86 | 109–111 +7–10 | 109–111 | 7–10 |

When this was fractionated on a column (1 cm × 112 cm) of Bio-Gel P-10 in 0.1 M-acetic acid, two cysteine-containing peptide fractions were obtained, one which was eluted in the void volume of the column and the other which corresponded to the peptides linked by the Cys-19–Cys-85 disulphide bridge. The fraction eluted in the void volume was freeze-dried and oxidized with performic acid at −10°C. The amino acid composition of this oxidized material is given in Table 5. The oxidized material was dissolved in 0.5 M-acetic acid (100 $\mu$l) containing 0.4% Brij 96 and fractionated on the same column as above. Two major peptide fractions were obtained; one was eluted in the void volume, the other at 70.5 ml and these had amino acid compositions (Table 5) consistent with their being derived from residues 109–111 and 7–10 respectively of the intact Thy-1 glycoprotein (Fig. 3). These results thus establish that the half-cystine residue at position 111 is linked to the half-cystine residue at position 9.

The peptide with Cys-111 that was eluted in the void volume was also sequenced by automated Edman degradation and gave an *N*-terminal sequence Val-Lys-, which agrees with the sequence for residues 109 and 110 as found in the sequencer run on T-10 and SP-6 (Fig. 3). The peptide also showed the unknown peak on amino acid analysis running between lysine and $NH_3$, as did T-10 and SP-6. The area of this unidentified ninhydrin positive peak was 1.6:1 compared with lysine, which is similar to the value for T-10 and SP-6.

## Discussion

With the exception of *C*-terminal peptides the sequence of the Thy-1 glycoprotein was obtained in a conventional manner by sequencing the tryptic and V-8 proteinase peptides. In this way a sequence to the end of peptide T-9 (residue 105) was originally obtained and parts of this were referred to in a preliminary publication concerning homologies between Thy-1 glycoprotein and immunoglobulin sequences (Campbell *et al.*, 1979). In this partial sequence residue 90 was referred to as glutamate but designated uncertain. It is now unequivocally established that this is serine, since no glutamate was detected at all in two sequencer runs on SP-6. The identification of glutamate came from a sequencer

run on T-9 and was probably due to contamination of T-9 with T-4 and to the fact that the serine thiazolinone derivative is unstable and is thus obtained in poor yield after high-pressure liquid chromatography. There is now no evidence for more than one amino acid at any position in the sequence and the fact that brain Thy-1 glycoprotein runs as a doublet on polyacrylamide gels in SDS is presumably due to heterogeneity in the carbohydrate.

Because the *C*-terminal peptides (apart from SP-6) were routinely lost (unless detergent was added), it was thought initially that reduced and alkylated Thy-1 antigen contained only three *S*-carboxymethylcysteine residues and this was not disproven by an initial sequencer run on SP-6 which did not extend to the end of the peptide. The initial amino acid analyses (Barclay *et al.*, 1976) were also not conclusive since they gave 3.4 cysteine residues per 110 amino acids. The existence of Cys-111 became clear in a second sequencer run on SP-6 and when the disulphide bonds were determined. Also further amino acid analyses on reduced and alkylated Thy-1 gave 4.3 *S*-carboxymethylcysteine residues/molecule (Table 4) and an analysis of oxidised Thy-1 gave 3.9 cysteic acid residues/molecule. The sequence of 111 residues fulfils the composition in Table 4, with the possibility that one residue each of alanine, glycine and glutamate is unaccounted for. However, it is not obvious how any extra residues can be accommodated in the sequence and these minor discrepancies are presumably due to the artifacts.

*Carbohydrate attachment points*

The carbohydrate composition (Barclay *et al.*, 1976) suggests the presence of carbohydrate structures *N*-linked to asparagine, with the possibility of another structure due to the presence of small amounts of galactosamine. Three *N*-linked structures were determined on the basis that the relevant peptides contained glucosamine and that at one position no residue was obtained in the sequencer and an Asx residue in the composition was not accounted for. The carbohydrate attachment sequences were Asn(23)-Asn-Thr; Asn(74)-Phe-Thr; Asn(98)-Lys-Thr. In each case these show an Asn-Xaa-Thr sequence and this sequence (or Asn-Xaa-Ser) is believed to be essential but not sufficient for *N*-linkage of carbohydrate (Neuberger *et al.*, 1972). One other sequence [Asn(93)-Pro-Thr] also showed this pattern but was unglycosylated. This is consistent with results *in vitro* which showed that peptides with an Asn-Pro-Thr sequence were not *N*-glycosylated (Bause & Hettkamp, 1979).

Galactosamine and glucosamine were detected in peptides SP-6 and T-10. It would normally be expected that galactosamine would be *O*-linked to serine or threonine and glucosamine to asparagine (Kornfeld & Kornfeld, 1976) but this cannot be the case for peptide T-10 which lacks these residues.

*Membrane orientation of Thy-1 glycoprotein and the C-terminus*

Much of the sequence and all of the carbohydrate of a number of cell surface molecules is exposed on the outside of the cell surface with a small sequence of hydrophobic amino acids at the *C*-terminus functioning to integrate the molecule into the lipid bilayer. A model for this mode of membrane integration is the major human erythrocyte sialoglycoprotein which spans the membrane via a sequence of 23 hydrophobic amino acids (Tomita & Marchesi, 1975). The HLA, A, B antigens also follow this pattern (Springer & Strominger, 1976) as does the influenza haemagglutinin (Skehel & Waterfield, 1975). Cytochrome *b*5 also has a hydrophobic amino acid sequence, but does not appear to span the lipid bilayer (Ozols & Gerard, 1977).

The Thy-1 glycoprotein has the properties of a molecule with a hydrophobic portion (see the introduction), yet it is very difficult to see how any of the sequence could be found in the membrane or inside the cell. If it is accepted that cystine residues and carbohydrate will be outside the cell, then this means that out of 111 residues the positions 9, 19, 23, 74, 85, 98 and 111 must be external to the cell. Furthermore, it seems unlikely that the sequence 24–73 could form a hydrophobic segment, since acidic or basic amino acids occur at regular intervals.

Despite the lack of an obvious hydrophobic sequence, the *C*-terminal peptides have hydrophobic properties. Without detergents they were either lost or chromatographed as aggregates as seen with SP-6. With Brij 96 the tryptic and peptic peptides were recovered apparently bound to the detergent micelles. The most striking example of this was the observation that the *C*-terminal tripeptide Val-Lys-Cys(111), obtained after pepsin digest and oxidation, was eluted from Bio-Gel P-10 with Brij 96 micelles. Other unusual features of the *C*-terminal peptides were the presence of unidentified ninhydrin-positive material and also of galactosamine and glucosamine (see above).

The simplest explanation for the hydrophobic behaviour of the *C*-terminal peptides would be that they have covalently coupled to them a hydrophobic group which is not protein. It could be that lipid is attached to the Thy-1 *C*-terminus, and there are precedents for the coupling of lipids to protein. This occurs through the sulphydryl and amino groups of the *N*-terminal cysteine of *Escherichia coli* murein-lipoprotein (Hantke & Braun, 1973) and through the sulphydryl of the *C*-terminal cysteine of rhodotorucine A of *Rhodosporidium toruloides* (Kamiya *et al.*, 1979). Also a glycoprotein of

vesicular stomatitis virus appears to have fatty acid attached to serine or threonine (Schmidt & Schlesinger, 1979). The same type of ester linkage has been postulated in a membrane protein from myelin of both human and bovine origin (Gagnon *et al.*, 1971; Stoffyn & Folch-Pi, 1971) as well as in a protein constituent of the sarcoplasmic reticulum membrane of rabbit skeletal muscle (MacLennan *et al.*, 1972). With Thy-1, lipid could not be attached through the sulphydryl of Cys-111 (which is bonded to Cys-9) or through the ε-amino group of Lys-110 (this can be succinylated). This leaves the carboxyl of Cys-111 as the likely attachment site of any other material. Further studies on the *C*-terminal peptides will be needed to determine their chemical constituents. Until this is done, the identification of Cys-111 as the *C*-terminal residue remains tentative.

*Sequence and structural homology with immunoglobulin*

A surprising result from the sequence of the Thy-1 glycoprotein was the finding of apparent sequence and structural homology with immunoglobulin. This was suggested in a preliminary report (Campbell *et al.*, 1979) and the determination of the disulphide bonds supports the contention that the polypeptide antigen may be folded in the manner characteristic of immunoglobulin. The homologies between immunoglobulin and Thy-1 glycoprotein are analysed in detail elsewhere (Cohen *et al.*, 1980).

We thank Gaynor Newton, Tony Gascoyne and Tony Willis for technical assistance and Jill Kosak for typing the manuscript.

## References


Acton, R. T., Morris, R. J. & Williams, A. F. (1974) *Eur. J. Immunol.* **4**, 598–602

Allen, A. K. & Neuberger, A. (1975) *FEBS Lett.* **60**, 76–80

Arndt, R., Stark, R. & Thiele, H.-G. (1977) *Immunology* **33**, 101–107

Barclay, A. N. & Hydén, H. (1978) *J. Neurochem.* **31**, 1375–1391

Barclay, A. N., Letarte-Muirhead, M. & Williams, A. F. (1975) *Biochem. J.* **151**, 699–706

Barclay, A. N., Letarte-Muirhead, M., Williams, A. F. & Faulkes, R. (1976) *Nature (London)* **263**, 563–567

Bause, E. & Hettkamp, H. (1979) *FEBS Lett.* **108**, 341–344

Beavan, G. H. & Holiday, E. R. (1952) *Adv. Protein Chem.* **7**, 319–386

Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029–3035

Bridgen, P. J., Cross, G. A. M. & Bridgen, J. (1976) *Nature (London)* **263**, 613–614

Campbell, D. G. (1978) D.Phil. Thesis, Oxford University

Campbell, D. G., Williams, A. F., Bayley, P. M. & Reid, K. B. M. (1979) *Nature (London)* **282**, 341–342

Cohen, F. E., Novotny, J., Sternberg, M. J. E., Campbell, D. G. & Williams, A. F. (1981) *Biochem. J.* **195**, 31–40

Dalchau, R. & Fabre, J. W. (1979) *J. Exp. Med.* **149**, 576–591

Douglas, T. C. (1972) *J. Exp. Med.* **136**, 1054–1062

Fields, K. L., Brockes, J. P., Mirsky, R. & Wendon, L. M. B. (1978) *Cell* **14**, 43–51

Gagnon, J., Finch, P. R., Wood, D. D. & Moscarello, M. A. (1971) *Biochemistry* **10**, 4756–4763

Goldschneider, I., Gordon, L. K. & Morris, R. J. (1978) *J. Exp. Med.* **148**, 1351–1366

Hantke, K. & Braun, V. (1973) *Eur. J. Biochem.* **34**, 284–296

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133

Hunt, S. V., Mason, D. W. & Williams, A. F. (1977) *Eur. J. Immunol.* **7**, 817–823

Kamiya, Y., Sakurai, A., Tamura, S., Takahashi, N., Tsuchiya, E., Abe, K. & Fukui, S. (1979) *Agric. Biol. Chem.* **43**, 363–369

Katz, A. M., Dreyer, W. J. & Anfinsen, C. B. (1959) *J. Biol. Chem.* **234**, 2897–2900

Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131

Kornfeld, R. & Kornfeld, S. (1976) *Annu. Rev. Biochem.* **45**, 217–237

Kuchel, P. W., Campbell, D. G., Barclay, A. N. & Williams, A. F. (1978) *Biochem. J.* **169**, 411–417

Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685

Ledbetter, J. A. & Herzenberg, L. A. (1979) *Immunol. Rev.* **47**, 63–90

Lennon, V. A., Unger, M. & Dulbecco, R. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6093–6097

Lesley, J. F. & Lennon, V. A. (1977) *Nature (London)* **268**, 163–165

Letarte, M. & Meghji, G. (1978) *J. Immunol.* **121**, 1718–1725

Letarte-Muirhead, M., Barclay, A. N. & Williams, A. F. (1975) *Biochem. J.* **151**, 685–697

Mason, D. W. & Williams, A. F. (1980) *Biochem. J.* **187**, 1–20

McClain, L. D., Tomana, M. & Acton, R. T. (1978) *Brain Res.* **159**, 161–171

MacLennan, D. H., Yip, C. C., Iles, G. H. & Seeman, P. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 469–477

Mendez, E. & Lai, C. Y. (1975) *Anal. Biochem.* **68**, 47–53

Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. (1972) in *Glycoproteins* (Gottschalk, A., ed.) 2nd edn., vol. 5, part A, pp. 450–490, BBA Library, Elsevier

Owen, M. J., Knott, J. C. A. & Crumpton, M. J. (1980) *Biochemistry* **19**, 3092–3099

Ozols, J. & Gerard, C. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3725–3729

Pan, S. C. & Dutcher, J. D. (1956) *Anal. Chem.* **28**, 836–838

Porath, J. (1974) *Methods Enzymol.* **34**, 13–29

Pruss, R. M. (1979) *Nature (London)* **280**, 688–689

Raff, M. C. (1971) *Transpl. Rev.* **6**, 52–80

Reif, A. F. & Allen, J. M. V. (1964) *J. Exp. Med.* **120**, 413–433
Ritter, M. A. & Morris, R. J. (1980) *Immunology* **39**, 85–91
Scheid, M., Boyse, E. A., Carswell, E. A. & Old, L. J. (1972) *J. Exp. Med.* **135**, 938–955
Schmidt, M. F. G. & Schlesinger, M. J. (1979) *Cell* **17**, 813–819
Skehel, J. J. & Waterfield, M. D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 93–97
Springer, T. A. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2481–2485
Stern, P. L. (1973) *Nature (London) New Biol.* **246**, 76–78
Stoffyn, P. & Folch-Pi, J. (1971) *Biochem. Biophys. Res. Commun.* **44**, 157–161
Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624–628
Tarr, G. E. (1978) *Methods Enzymol.* **47**, 335–357
Thierfelder, S. (1977) *Nature (London)* **269**, 691–693
Tomita, M. & Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2964–2968
Trowbridge, I. S., Weissman, I. L. & Bevan, M. J. (1975) *Nature (London)* **256**, 652–654
Williams, A. F. (1976) *Eur. J. Immunol.* **6**, 526–528
Williams, A. F., Barclay, A. N., Letarte-Muirhead, M. & Morris, R. J. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 51–61
Woods, K. R. & Wang, K.-T. (1967) *Biochim. Biophys. Acta* **133**, 369–370
Zwerner, R. K., Barstad, P. A. & Acton, R. T. (1977) *J. Exp. Med.* **146**, 986–1000