

Factor \bar{D} of the Alternative Pathway of Human Complement

PURIFICATION, ALIGNMENT AND *N*-TERMINAL AMINO ACID SEQUENCES OF THE MAJOR CYANOGEN BROMIDE FRAGMENTS, AND LOCALIZATION OF THE SERINE RESIDUE AT THE ACTIVE SITE

Diana M. A. JOHNSON, Jean GAGNON and Kenneth B. M. REID

Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

(Received 20 November 1979)

1. The serine esterase factor \bar{D} of the complement system was purified from outdated human plasma with a yield of 20% of the initial haemolytic activity found in serum. This represented an approx. 60 000-fold purification. 2. The final product was homogeneous as judged by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (with an apparent mol.wt. of 24 000), its migration as a single component in a variety of fractionation procedures based on size and charge, and its *N*-terminal amino-acid-sequence analysis. 3. The *N*-terminal amino acid sequence of the first 36 residues of the intact molecule was found to be homologous with the *N*-terminal amino acid sequences of the catalytic chains of other serine esterases. Factor \bar{D} showed an especially strong homology (>60% identity) with rat 'group-specific protease' [Woodbury, Katunuma, Kobayashi, Titani, & Neurath (1978) *Biochemistry* 17, 811–819] over the first 16 amino acid residues. This similarity is of interest since it is considered that both enzymes may be synthesized in their active, rather than zymogen, forms. 4. The three major CNBr fragments of factor \bar{D} , which had apparent mol.wts. of 15 800, 6600 and 1700, were purified and then aligned by *N*-terminal amino acid sequence analysis and amino acid analysis. By using factor \bar{D} labelled with di-[1,3- 14 C]isopropylphosphofluoridate it was shown that the CNBr fragment of apparent mol.wt. 6600, which is located in the C-terminal region of factor \bar{D} , contained the active serine residue. The amino acid sequence around this residue was determined.

Factor \bar{D} is an enzyme of the alternative pathway of complement that has some of the characteristics of a typical serine esterase. It is composed of a catalytic chain of approx. 23 000–25 000 mol.wt. (Götze, 1976; Volanakis *et al.*, 1977) and is irreversibly inhibited by iPr_2P-F (Fearon *et al.*, 1974). It has an *N*-terminal amino acid sequence of Ile-Leu-Gly-Gly- (Volanakis *et al.*, 1977; Davis *et al.*, 1979a,b) which is very characteristic of the type of highly conserved amino-acid sequences found at the *N*-terminal end of the catalytic chains of other serine esterases (de Haën *et al.*, 1975). It is not known whether factor \bar{D} is synthesized first as a zymogen or in its active form. There is one report that a small proportion of the enzyme present in the

plasma is in the proenzyme form and that conversion of this material to the active enzyme can be achieved by trypsin treatment (Fearon *et al.*, 1974). However recent studies suggest that it is probable that factor \bar{D} may normally be present in plasma only in its activated form (Lesavre & Müller-Eberhard, 1978). Factor B is another serine-esterase component of the alternative pathway that has clearly been shown to be present in plasma in its zymogen form. When factor B is complexed with C3b, in the presence of Mg^{2+} , it is readily split by factor \bar{D} at a single Arg–Lys bond thus yielding the Ba and Bb fragments (Lesavre *et al.*, 1979). This limited proteolysis of Factor B brings about the formation of the active enzyme complex $\bar{C}3b$ Bb which can split and activate C3, the most abundant complement component, which plays a central role in the activation of the complement system (Schreiber *et al.*, 1978; Austen, 1978). The initial activation steps of the alternative pathway are not

Abbreviations used: SDS, sodium dodecyl sulphate; iPr_2P-F , di-isopropyl phosphorofluoridate; Quadrol, *NNN'*-tetrakis-(2-hydroxypropyl)ethenediamine trifluoroacetate; Polybrene, 1,5-diaza-1,5-dimethylundecamethylene polymethobromide.

fully understood and structural studies may help clarify the role that factor D plays in these steps.

The study of factor D̄ has been hampered by its low serum concentration (approx. 1–2 mg/litre) and by the difficulty in removing contaminants of similar apparent molecular weight and behaviour on ion-exchange columns. This paper describes the isolation of highly purified factor D̄ in sufficient quantities to carry out *N*-terminal sequence analysis on the intact molecule, and purification and *N*-terminal sequence analysis of the peptides generated by CNBr treatment of the reduced and alkylated intact molecule.

The results obtained confirm, and clarify, the *N*-terminal sequence obtained for the first nine residues reported by Davis *et al.* (1979a,b) and extend the *N*-terminal sequence information to residue 36. Amino-acid analysis and amino-acid sequencing has also allowed the alignment of the three major peptides produced by CNBr treatment. Labelling of factor D̄ with [1,3-¹⁴C]iPr₂P-F has allowed the identification and localization of the serine residue at the active site.

Materials and Methods

Materials

Fresh frozen human plasma and outdated human plasma, which had been collected in sodium citrate anticoagulant, was obtained from the Churchill Hospital, Oxford. Plasma was clotted overnight, at 4°C, by the addition of 1.0M-CaCl₂ to give a final concentration of 20mM. The clot was removed by centrifugation and filtration through muslin, and the serum was stored at -20°C.

Fresh frozen plasma was clotted and used for the preparation of factor D̄-deficient serum as described by Martin *et al.* (1976), using a Sephadex G-75 column equilibrated with 0.1M-Tris/HCl/0.2M-NaCl/2mM-EDTA, pH 8.0 at 4°C.

CM-Sephadex C-50, Sephadex G-75, concanavalin A-Sepharose, Sephadex G-50 (superfine) and Sephadex G-75 (superfine) were all obtained from Pharmacia Fine Chemicals, Uppsala, Sweden. CM-cellulose 32 was obtained from Whatman, Maidstone, Kent, U.K.

Ovalbumin, bovine α-chymotrypsinogen A, whale skeletal muscle myoglobin, egg white lysozyme, horse-heart cytochrome *c* and bovine insulin were all obtained from Sigma. Bovine serum albumin was obtained from Armour Pharmaceutical Co., Eastbourne, Sussex, U.K. Iodoacetamide, iodoacetic acid and dithiothreitol were obtained from BDH. CNBr was from Koch-Light Laboratories, Colnbrook, Bucks., U.K.

Quadrol was purchased from Beckman Instruments, Palo Alto, CA, U.S.A. Polybrene was from Aldrich Chemical Co., R.N. Emanuel, Wembley,

Middlesex, U.K. All other chemicals used in the sequencer were obtained from Rathburn Chemicals, Walkerburn, Peebleshire, Scotland, U.K.

Iodo[2-¹⁴C]acetic acid (54 Ci/mol) was from The Radiochemical Centre, Amersham, Bucks., U.K. [1,3-¹⁴C]iPr₂P-F (100 Ci/mol) was purchased from New England Nuclear.

Haemolytic assay for factor D̄

Factor D̄ was assayed using the haemolytic diffusion plate assay of Martin *et al.* (1976). The assay was standardized by using dilutions of pooled normal serum. One unit of factor D̄ activity was defined as the haemolytic area obtained using 10 μl of the standard serum.

Purification of factor D̄

Human serum (2000 ml) was dialysed for 20 h at 4°C against 20 litres of 5 mM-EDTA, pH 5.4. The euglobulin precipitate was removed by centrifugation at 2000g for 1 h. The conductivity of the pseudoglobulin solution was adjusted to 12–14 mS using 1M-sodium phosphate buffer, pH 6.0, and the solution was loaded on to a column (12 cm × 8 cm) of CM-Sephadex C-50, equilibrated in 0.2M-sodium phosphate buffer, pH 6.0 (6.15 ml of 0.2M-Na₂HPO₄ + 43.85 ml of 0.2M-NaH₂PO₄/50 ml of buffer). The column was washed with 1500 ml of 0.2M-sodium phosphate, pH 6.0, and then the factor D̄ was eluted using 2000 ml of 0.4M-sodium phosphate buffer, pH 6.0 (12.3 ml of 0.4M-Na₂HPO₄ + 87.7 ml of 0.4M-NaH₂PO₄/100 ml of buffer). The flow rate was 600 ml/h during application of the 0.2M-sodium phosphate wash, and 50 ml/h during the elution of factor D̄.

The fractions containing factor D̄ haemolytic activity were pooled and (NH₄)₂SO₄ was added to 50% saturation (291 g/litre). The suspension was stirred for 2 h at 4°C, then centrifuged at 10000g for 2 h. (NH₄)₂SO₄ was added to the supernatant (125 g/litre) to give 70% saturation and the suspension was stirred for 2 h at room temperature. The precipitate was dissolved in 40 ml of 0.1M-Tris/HCl/0.2M-NaCl/2mM-EDTA, pH 8.0 at 4°C, and, after centrifugation, applied to a column (5 cm × 100 cm) of Sephadex G-75 equilibrated with the same buffer. The column was run at a flow rate of 30 ml/h. The fractions containing factor D̄ haemolytic activity were pooled and concentrated to 10 ml using an Amicon Diaflo Ultrafiltration cell fitted with a UM2 membrane.

These initial steps in the purification procedure were developed by R. Prohaska and provide a factor D̄ preparation of sufficient purity for functional studies. However, for structural studies, further purification of the factor D̄ was considered necessary.

The material from the Sephadex G-75 steps from four different preparations was pooled and dialysed against 0.23 M-acetic acid/NaOH buffer, pH 5.2 (5.5 ml of 4 M-NaOH adjusted to pH 5.2 with acetic acid and made up to 100 ml) and applied to a column (1.5 cm \times 20 cm) of CM-cellulose 32 that had been equilibrated with the same buffer. The column was washed with the starting buffer at 30 ml/h until the A_{230} of the eluate was zero, and then a linear gradient was developed using 120 ml of 0.23 M-acetic acid/NaOH buffer, pH 5.2, and 120 ml of the same buffer containing 0.3 M-NaCl. The fractions containing factor \bar{D} haemolytic activity were pooled and concentrated to 10 ml, then dialysed against 0.1 M-sodium acetate/acetic acid buffer (0.1 M-sodium acetate adjusted to pH 6.0 with acetic acid) containing 1 M-NaCl, 1 mM-MgCl₂, 1 mM-MnCl₂ and 1 mM-CaCl₂. After dialysis the sample (10 ml) was applied to a column (1 cm \times 12 cm) of concanavalin A-Sepharose, equilibrated with the same buffer, and eluted at a flow rate of 10 ml/h. The factor \bar{D} haemolytic activity was not retarded on the column. A contaminant, of the same apparent molecular weight as factor \bar{D} in non-reducing conditions on SDS/polyacrylamide gels, was bound to the column and could be eluted with 2.5% (w/v) 1-O-methyl- α -D-glucopyranoside dissolved in the same buffer used to equilibrate the column.

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

SDS/polyacrylamide gel electrophoresis and determination of the apparent molecular weights of factor \bar{D} and its CNBr fragments

Electrophoresis was carried out as described by Laemmli (1970) using 15% or 20% (w/v) polyacrylamide slab gels containing 1% (w/v) SDS. Proteins were stained using Coomassie Brilliant Blue.

Non-reduced protein samples (5–10 μ g) in 20 mM-iodoacetamide/0.1 M-Tris/HCl/4 M-urea/1% (w/v) SDS, pH 8.0, were incubated for 2 min at 100°C. Reduced samples were prepared by incubation with 20 mM-dithiothreitol for 2 min at 100°C.

The apparent molecular weights of factor \bar{D} and its CNBr fragments were calculated by comparison with the mobilities of reduced and non-reduced marker proteins and peptides: bovine serum albumin (mol.wt. 67 000), ovalbumin (mol.wt. 42 000), bovine pancreas α -chymotrypsinogen A (mol.wt. 25 700), whale skeletal muscle myoglobin (mol.wt. 17 200), egg-white lysozyme (mol.wt. 14 300), horse-heart cytochrome *c* (mol.wt. 12 400), bovine insulin (mol.wt. 5800), B chain of insulin (mol.wt. 3400) and A chain of insulin (mol.wt. 2500).

Reduction and alkylation of factor \bar{D}

Factor \bar{D} (90 nmol) was dissolved in 6 M-guanidine/HCl/0.4 M-Tris/HCl/2 mM-EDTA, pH 8.0 (1.5 ml) containing 20 mM-dithiothreitol, and incubated in a sealed flask at 37°C for 3 h with frequent mixing. Iodo[2-¹⁴C]acetic acid (100 μ Ci in 100 μ l of distilled water) was added and the incubation was continued for 5 min at 25°C; then ice-cold carrier iodoacetic acid was added to a concentration of 60 mM, and the mixture was cooled on ice for 45 min. The reduced and alkylated sample was then dialysed at 4°C against distilled water (1 litre) for 3 h, followed by two changes of 0.5 M-acetic acid (1 litre) and was freeze-dried.

Labelling of factor \bar{D} with [1,3-¹⁴C]iPr₂P-F

Factor \bar{D} (50 nmol) in 5 ml of 0.01 M-Tris/HCl/0.15 M-NaCl/1 mM-MgCl₂/1 mM-CaCl₂, pH 8.0 at 4°C, was incubated at 30°C for 1 h with 1 μ mol of [1,3-¹⁴C]iPr₂P-F (100 μ Ci). Non-radioactive iPr₂P-F was added to give a final concentration of 10 mM and the mixture was incubated for a further 1 h at 30°C. The inactivated factor \bar{D} was dialysed for 24 h against three changes of 0.5 M-acetic acid and was freeze-dried. Radioactive samples [in 10–30 μ l of 5% (v/v) formic acid] were added to 10 ml of 1,4-dioxan containing 2% (w/v) naphthalene and 0.5% (w/v) 2,5-diphenyloxazole. Samples were counted for radioactivity using an LKB-Wallac 1210 Ultrobeta counter.

CNBr digestion of factor \bar{D} and purification of the major CNBr peptides

Reduced and alkylated factor \bar{D} (50 nmol) was dissolved in 70% (v/v) formic acid (0.7 ml) containing CNBr (0.15 mmol) and incubated for 16 h at 25°C. The mixture was freeze-dried and then redissolved in 70% (v/v) formic acid (0.7 ml) and applied to a column (1.5 cm \times 100 cm) of Sephadex G-50 (superfine) equilibrated with 5% (v/v) formic acid. Peptides eluting near the void volume of the Sephadex G-50 column were further purified by gel filtration on a column (1.5 cm \times 100 cm) of Sephadex G-75 (superfine) equilibrated with 10% (v/v) formic acid. Smaller peptides were further purified by high pressure liquid chromatography using a μ Bondapak C-18 column. Two solutions were used in the high pressure liquid chromatography of peptides: solution A, which was composed of CH₃CN and 10 mM-NH₄HCO₃ (5:95, v/v) and solution B, which was composed of CH₃CN and 10 mM-NH₄HCO₃ (64:40, v/v). The column was equilibrated with a mixture of solution A and solution B in the ratio 95:5 (v/v) and then eluted with a linear gradient to give a final ratio of 10:90 (v/v) of solution A to solution B. The column was pumped at 1.5 ml/min by a Waters Associates system (model 6000 A).

Amino acid analysis

Proteins and peptides were hydrolysed, under vacuum, at 110°C for 48 h in constant-boiling HCl containing 4 mM-phenol. No corrections were made for destruction of threonine and serine, and tryptophan was not measured. Cystine was measured as *S*-carboxymethylcysteine after acid hydrolysis, under vacuum, of reduced and alkylated samples in the presence of 0.05% (v/v) 2-mercaptoethanol.

Automated amino-acid sequence determination

The *N*-terminal amino acid sequences of factor \bar{D} and its CNBr fragments were determined by automated Edman degradation in a Beckman 890C sequencer using the 0.3 M-Quadrol programme of Hunkapiller & Hood (1978). Polybrene (2 mg) was added to the sequencer cup prior to application of each sample (Klapper *et al.*, 1978). The thiazolinones released were converted into the amino-acid phenylthiohydantoin derivatives by heating at 80°C under N₂ for 10 min in 1 M-HCl (200 μ l) containing 0.1% (v/v) ethanethiol. After two extractions with ethyl acetate (700 μ l), 40–80% of the extracted amino acid phenylthiohydantoin derivative was examined by high-pressure liquid chromatography. The chromatography was done on a μ Bondapak C-18 column using a 14–56% methanol gradient in aqueous buffer (Bridgen *et al.*, 1976) pumped at 2.2 ml/min by a Waters Associates system (model 6000 A) with a u.v. detector.

Results

Purification of factor \bar{D}

A summary of the purification of factor \bar{D} is given in Table 1. After precipitation of euglobulin, more than 99% of the protein remaining in the pseudoglobulin fraction was removed by ion-exchange chromatography on CM-Sephadex C-50. On washing the column with 0.4 M-sodium phosphate buffer, pH 6.0, the factor \bar{D} haemolytic activity

eluted above 16 mS. The activity peak was broad and no improvement in purification resulted from eluting factor \bar{D} with a linear gradient instead of the stepwise wash. It was found convenient to concentrate the material from the CM-Sephadex C-50 column by (NH₄)₂SO₄ fractionation. Most of the factor \bar{D} haemolytic activity precipitated in the 50–70% saturation fraction.

The (NH₄)₂SO₄ precipitate was further purified by gel filtration on Sephadex G-75, where the factor \bar{D} haemolytic activity was separated from a large number of proteins all having apparent molecular weights >30000. Because of the small amount of protein (usually 7–10 mg) in the factor \bar{D} pool from the Sephadex G-75 column it was found convenient to combine the Sephadex G-75 factor \bar{D} pools from four preparations before carrying out the final stages of the purification. The partially purified factor \bar{D} from Sephadex G-75 was therefore stored at –70°C until use.

Following gel filtration on Sephadex G-75, the preparation appeared to be functionally pure with respect to factor \bar{D} haemolytic activity, but on further purification on CM-cellulose 32 at least four protein peaks were eluted (Fig. 1*a* and *b*). The factor \bar{D} haemolytic activity eluted between 22–27 mS. Another protein, which was removed by use of concanavalin-A-Sepharose, was eluted between 25–28 mS in the gradient on CM-cellulose 32 (Fig. 1). This protein could not be distinguished from factor \bar{D} in non-reducing conditions on SDS/polyacrylamide gels, but in reducing conditions had a higher apparent molecular weight compared to factor \bar{D} (Fig. 1*b*). No factor \bar{D} haemolytic activity was associated with this contaminating protein, the concentration of which varied considerably from preparation to preparation, sometimes comprising approx. 50% of the protein in the partially purified factor \bar{D} pool from CM-cellulose 32. Thus the CM-cellulose 32 pool was routinely applied to a column of concanavalin-A-Sepharose. Factor \bar{D}

Table 1. Purification of factor \bar{D} from serum

Fraction	Volume (ml)	Total protein (A ₂₈₀)	10 ⁻³ × total haemolytic activity (units)	Specific activity (units/A ₂₈₀)	Yield of activity (%)	Purification factor
Serum	2000	110 600	191	1.73	100	1.0
Pseudoglobulin	2100	104 300	176	1.69	92	0.98
CM-Sephadex C-50	900	211	160	758	84	440
50–70% (NH ₄) ₂ SO ₄ precipitate	—	72.4	130	1790	68	1040
Sephadex G-75	350	8.52	86.5	10 200	45	5890
CM-cellulose-32	86	1.56	46.2	29 600	24	17 100
Concanavalin-A-Sepharose	24	0.875*	43.1	49 200	22	28 500

* The final yield of protein as estimated by amino-acid analysis was 0.42 mg. This gives a final specific activity (units/mg of protein) for the factor \bar{D} of 102 600 and a purification factor of approx. 60 000.

was not retained on this column, but the contaminant was bound and could only be eluted with 1-*O*-methyl- α -D-glucopyranoside.

The purified factor \bar{D} appeared to be homogeneous on SDS/polyacrylamide gels in reducing and non-reducing conditions (Fig. 2). On staining the gels with Coomassie Blue, the protein band appeared broad in the non-reduced sample, but amino-acid sequence analysis has shown the presence of only one sequence and the number of fragments obtained after CNBr treatment is also consistent with there

being only one type of polypeptide chain in the final preparation.

Stability of factor \bar{D}

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

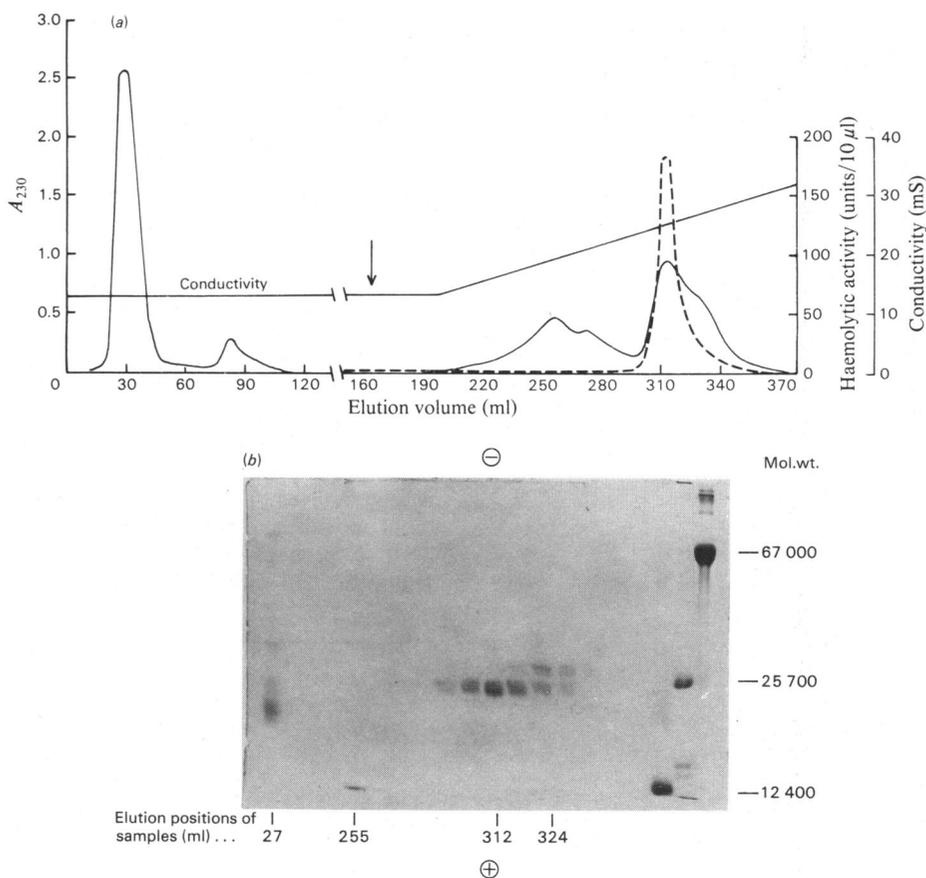


Fig. 1. Ion-exchange chromatography of factor \bar{D} on CM-cellulose 32, and SDS/polyacrylamide-gel electrophoresis (Laemmli, 1970) of the fractions

(a) The partially purified factor \bar{D} (approx. 30 mg) from Sephadex G-75 was equilibrated with 0.23 M-acetic acid/NaOH buffer, pH 5.2, and loaded on to a column (1.5 cm \times 20 cm) of CM-cellulose 32 equilibrated in the same buffer. The column was developed as described in the Materials and Methods section. Factor \bar{D} haemolytic activity eluted between 22–27 mS. —, A_{230} ; ---, haemolytic activity. The arrow marks the start of the gradient. (b) Samples (25 μ l) were taken from column fractions corresponding to elution at 27, 81, 210, 240, 255, 270, 285, 294, 300, 306, 312, 318, 324, 330, 345 and 360 ml on the chromatogram profile shown in (a). The samples were reduced and alkylated as described in the Methods and run on a SDS/15% (w/v) polyacrylamide slab gel. The major contaminants in the preparation were eluted at peak volumes of 27, 255 and 324 ml and factor \bar{D} was eluted at 312 ml. Albumin (mol.wt. 67 000), α -chymotrypsinogen A (mol.wt. 25 700) and myoglobin (mol.wt. 17 200) were run as marker proteins, after reduction and alkylation.

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

Molecular weight determinations

The apparent molecular weight of factor \bar{D} was determined on 15% (w/v) polyacrylamide gels containing 1% (w/v) SDS. In non-reducing conditions the molecular weight was determined as 24000 and in reducing conditions as 23700. Reduced and alkylated factor \bar{D} had a slightly slower mobility on SDS/polyacrylamide gels, compared with non-reduced factor \bar{D} (Fig. 2).

CNBr digestion of factor \bar{D}

From the methionine content of factor \bar{D} (1.2 methionine residues/100 residues) three, or possibly four, peptides were expected. The CNBr digest was initially fractionated on Sephadex G-50 (superfine) (Fig. 3). Two major peaks of protein, containing radioactivity, eluted at 56–76 ml and at 90–104 ml respectively. The second peak (90–104 ml) contained peptide CNBr-2. The first peak (56–76 ml) was found to be heterogeneous on SDS/20% (w/v) polyacrylamide gels and was repurified on a column (1.5 cm \times 100 cm) of Sephadex G-75 (super-

fine) equilibrated in 10% (v/v) aq. formic acid. Peptide CNBr-1 eluted between 53–63 ml on this column. Peptide CNBr-3 was purified from the pool made between 105–150 ml on the initial fractionation of the digest on the Sephadex G-50 (superfine) column (Fig. 3). This pool was freeze-dried and then redissolved in 10 mM- NH_4HCO_3 /5% (w/v) CH_3CN (50 μl) plus 99% formic acid (10 μl). The material in the pool was then further purified by high pressure liquid chromatography using a $\mu\text{Bondapak C-18}$ column. Two significant peaks of peptide material were detected in an A_{210} ratio of approx. 10:1. The amino-acid compositions of these peptides were identical, suggesting that they were two forms of the same peptide. It is possible that one of the peptides had its C-terminal homoserine in the lactone form.

The apparent molecular weights of peptides CNBr-1 and CNBr-2, as estimated on SDS/20% (w/v) polyacrylamide gels, were 15800 and 6600. Peptide CNBr-3 was not detected on SDS/20% (w/v) polyacrylamide gels, but its molecular weight was estimated to be 1700 from its amino-acid composition (Table 2) and its amino-acid sequence (Fig. 4). The combined molecular weights of these three major CNBr peptides thus account for the total apparent molecular weight of factor \bar{D} , although the possible occurrence of a fourth, very small, CNBr peptide has not been excluded.

Two other fragments were detected in CNBr digests of factor \bar{D} . One of these (of apparent mol.wt. 18000) eluted on Sephadex G-75 superfine just ahead of peptide CNBr-1, from which it was not clearly resolved. This peptide, which was present at about one-fifth of the concentration of peptide CNBr-1, probably resulted from incomplete digestion of the methionyl bond at position 15 in factor \bar{D}

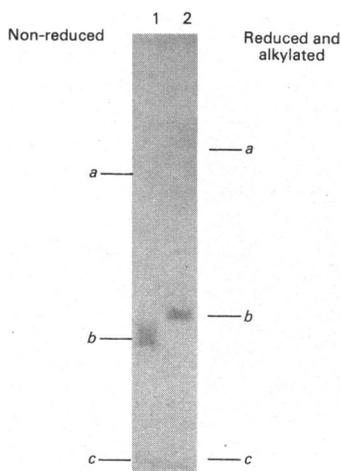


Fig. 2. SDS/polyacrylamide-gel electrophoresis (Laemmli, 1970) of purified factor \bar{D}

Factor \bar{D} (5–10 μg), after the final purification step on concanavalin-A-Sepharose, was reduced and alkylated as described in the Materials and Methods section. Non-reduced (track 1) and reduced (track 2) samples were run on a 15% (w/v) SDS/polyacrylamide slab gel. The arrows show the positions of the following marker proteins: (a) albumin (mol.wt. 67000), (b) -chymotrypsinogen A (mol.wt. 25700), (c) cytochrome *c* (mol.wt. 12400), before and after reduction and alkylation.

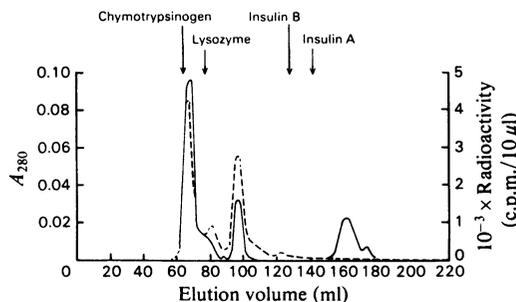


Fig. 3. Purification on Sephadex G-50 of peptides obtained by CNBr digestion of factor \bar{D}

The CNBr digest was freeze-dried, redissolved in 70% (v/v) formic acid (0.7 ml) and applied to a column (1.5 cm \times 100 cm) of Sephadex G-50 (superfine) equilibrated with 5% (v/v) formic acid and run at a flow rate of 2.5 ml/h. —, A_{280} ; ---, radioactivity of S -[^{14}C]carboxymethylcysteine.

(Fig. 4) and would be composed of peptides CNBr-3 and CNBr-1. The other fragment, also present in low amounts, had an apparent mol.wt. of 23000. This fragment was distinguishable from residual un-cleaved factor \bar{D} and is probably composed of peptides CNBr-1 and CNBr-2. Amino-acid sequence analysis on these two fragments was not performed.

Amino acid compositions, and N-terminal amino acid sequences, of factor \bar{D} and its CNBr fragments

The amino-acid compositions of factor \bar{D} , peptide CNBr-1, peptide CNBr-2 and peptide CNBr-3 are given in Table 2.

As expected from earlier work (Volanakis *et al.*, 1977) the *N*-terminal amino acid of factor \bar{D} was found to be isoleucine. The *N*-terminal amino-acid sequences of factor \bar{D} and peptide CNBr-3 were found to be identical for the first 11 residues, except for residue 10, which was not identified in the sequencer run of peptide CNBr-3 (Table 3). The amino acid at residue 10 was shown to be histidine in the sequencer run of intact factor \bar{D} and this was consistent with the presence of one histidine residue in the amino-acid composition of peptide CNBr-3 (Table 2).

Table 2. *Amino-acid compositions of factor \bar{D} and its major CNBr peptides*

The amino acid compositions are given as residues/100 residues except in the case of peptide CNBr-3 which is given as mol of residue/mol of peptide. The samples were hydrolysed at 110°C for 47h. No corrections were made for serine and threonine destruction, and tryptophan was not determined. Cystine was estimated as *S*-carboxymethylcysteine.

Amino acid	Amino acid composition of			
	Factor \bar{D} (residues/100 residues)	CNBr-1	CNBr-2	CNBr-3 (mol/mol)
Asp	9.4	10.6	10.0	0.3 (0)
Thr	3.5	3.7	3.5	0.3 (0)
Ser	5.0	5.4	7.6	—
Hse	—	0.6	—	0.7 (1)
Glu	9.4	9.8	7.9	2.1 (2)
Pro	8.1	7.3	6.3	1.1 (1)
Gly	10.4	10.2	14.3	2.2 (2)
Ala	8.8	10.6	9.0	2.7 (3)
Val	8.0	8.3	9.8	0.3 (0)
Cys	4.0	1.6	3.0	—
Met	1.2	—	—	—
Ile	2.6	2.2	4.3	0.8 (1)
Leu	10.3	14.1	6.8	1.0 (1)
Tyr	1.8	0.4	2.4	0.9 (1)
Phe	1.3	0.2	0.7	—
His	5.2	6.5	1.9	1.0 (1)
Lys	3.9	2.1	4.9	—
Arg	7.1	6.4	7.6	1.6 (2)

From the apparent mol.wt. (15800) of peptide CNBr-1 it was estimated that the peptide was approx. 141 amino acid residues long. The *N*-terminal amino acid of peptide CNBr-1 was found to be alanine. By comparing the *N*-terminal amino acid sequence of peptide CNBr-1 with that of intact factor \bar{D} (Table 3) it can be seen that the *N*-terminal amino acid sequence of peptide CNBr-1 overlaps with that of factor \bar{D} from residue 16 onward. The only exception was at residue 11 of peptide CNBr-1, where *S*-carboxymethylcysteine was found compared with an unidentified residue in factor \bar{D} (Table 3). However this would be expected since the intact factor \bar{D} used for amino acid sequence analysis was not reduced and alkylated.

From the sequence data obtained for peptides CNBr-1 and CNBr-3, and from the lack of homoserine in the amino acid composition of peptide CNBr-2, it is probable that peptide CNBr-2 is derived from the *C*-terminal portion of factor \bar{D} (Fig. 4). Peptide CNBr-2 was estimated to be composed of approx. 60 amino acid residues, and automated *N*-terminal sequence analysis gave sequence of the first 26 amino acid residues (with the exception of positions 21 and 22). When peptide CNBr-2 was isolated from a CNBr digest of factor \bar{D} that had been labelled with [1,3-¹⁴C]Pr₂P-F, it was found to be radiolabelled. By scintillation counting of one-third of the butyl chloride extracts from the sequencer run of peptide CNBr-2 it was shown that the radioactivity in the peptide was located in the serine residue found at position 14 (Table 3).

Discussion

The purification procedure described in this paper gave a final yield of about 0.2 mg of factor \bar{D} /1000 ml of serum, and a recovery of about 20% of the initial factor \bar{D} haemolytic activity (Table 1). The overall purification from serum was approx. 60000-fold. Addition of protease inhibitors in the early stages of the purification did not appear to increase the recovery of factor \bar{D} . The yield of factor \bar{D} is consistent with a serum concentration of approx. 1.0–1.5 μ g/ml, which is similar to the estimates of Lesavre *et al.* (1979) but slightly lower than suggested by an earlier report (Lesavre & Müller-Eberhard, 1978).

The estimate of an apparent mol.wt. of 24000 is similar to values obtained by others (Volanakis *et al.*, 1977). Factor \bar{D} that had been reduced and alkylated showed a slightly slower mobility on SDS/polyacrylamide gels compared with non-reduced factor \bar{D} (Fig. 2). This has also been observed by Volanakis *et al.* (1977). Lesavre *et al.* (1979) and Davis *et al.* (1979b) found that factor \bar{D} stained with Schiff stain after periodate oxidation, which indicates that it is a glycoprotein. In the

Table 3. Automated sequence analysis of factor \bar{D} and peptides CNBr-1, CNBr-2 and CNBr-3

Details of the automated Edman degradation are given in the text. Amino-acid phenylthiohydantoin derivatives were identified by high pressure liquid chromatography. The presence of the *S*-carboxymethylcysteine phenylthiohydantoin derivatives were also confirmed by detection of radioactive label. —, not identified.

Residue number	Amino acid identified in			
	Factor \bar{D}	Peptide CNBr-1	Peptide CNBr-2	Peptide CNBr-3
1	Ile	Ala	Cys	Ile
2	Leu	Ser	Ala	Leu
3	Gly	Val	Glu	Gly
4	Gly	Gln	Ser	Gly
5	Arg	Leu	Asn	Arg
6	Glu	Asn	Arg	Glu
7	Ala	—	Arg	Ala
8	Glu	Ala	Asp	Glu
9	Ala	—	Ser	Ala
10	His	Leu	Cys	—
11	Ala	Cys	Lys	—
12	Arg	Gly	Gly	Ala
13	Pro	Gly	Asp	—
14	Tyr	Val	Ser*	—
15	Met	Leu	Gly	—
16	Ala	Val	Gly	—
17	Ser	Ala	Pro	—
18	Val	Glu	Leu	—
19	Gln	—	Val	—
20	Leu	—	Cys	—
21	Asn	Val	—	—
22	—	Val	—	—
23	Ala	—	Val	—
24	Glu	—	Leu	—
25	Leu	—	Glu	—
26	—	—	(Gly)	—
27	Gly	—	—	—
28	Gly	—	—	—
29	Val	—	—	—
30	Leu	—	—	—
31	Val	—	—	—
32	Ala	—	—	—
33	Glu	—	—	—
34	Gln	—	—	—
35	—	—	—	—
36	Val	—	—	—
Amount used in sequencer (nmol)	60	20	30	12
Recovery at first step (nmol)	6.5	3.1	10.5	3.1
Stepwise yield (%)	93	90	94	75
Residues used to calculate yield	Leu-2,20,25,30	Leu-5,10,15	Cys-1,10,20	Ala-7,9,11

* Radioactivity was released at this position when peptide CNBr-2 was isolated from [1,3- ^{14}C]iPr $_2$ P-F-labelled factor \bar{D} .

purification procedure used in this study, however, it was observed that factor \bar{D} was not retained by a concanavalin A-Sepharose column.

The amino-acid composition obtained for factor \bar{D} shows no unusual features and is in good agreement with that reported by Davis *et al.* (1979b) (Table 2). However, the one other published amino-acid composition given for human factor \bar{D} (Volana-

kis *et al.*, 1977) differs in some aspects from the values given in this study (Table 2), e.g. isoleucine was reported to be present as 6.0 residues/100 residues compared with 2.6 residues/100 residues found in this study; alanine 5.8 compared with 8.8; cystine 2.9 compared with 4.0; leucine 6.9 compared with 10.3; lysine 7.1 compared with 3.9. The methionine value of 1.2 residues/100 residues

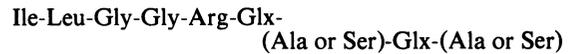
obtained in this study (Table 2) indicated that there were 2–3 methionine residues/molecule of factor \bar{D} , which should therefore yield three or four peptides on treatment with CNBr.

Digestion of factor \bar{D} with CNBr gave three major peptides, CNBr-1, CNBr-2 and CNBr-3 (Table 2 and Fig. 4) in yields of 26%, 55% and 39% respectively. The yields of peptides CNBr-1 and CNBr-3 were lower than that of peptide CNBr-2 since they each underwent an extra purification step compared with peptide CNBr-2. The amino-acid-sequence data given allows the alignment of these CNBr peptides as shown in Fig. 4. The possibility that there may be a fourth, very small, peptide has not been eliminated and must be considered, especially as it can be estimated that intact factor \bar{D} contains approx. 2.5 residues of methionine/molecule (Table 2). That any unidentified peptide must be small is suggested from the finding that the apparent molecular weight of factor \bar{D} is virtually accounted for by the sum of the molecular weights of peptides CNBr-1, CNBr-2 and CNBr-3 (15800, 6600 and 1700 respectively).

The yield of *N*-terminal phenylthiohydantoin amino acid at the first cycle in each of the sequencer runs performed on factor \bar{D} and the CNBr fragments was usually low (Table 3). The reasons for this low yield in most of the sequencer runs are not known, but could involve losses of the polypeptide from the cup during the washing steps in the first cycle of the sequencer programme. The possibility that general technical difficulties caused the low yields can be excluded since myoglobin standards, run on the sequencer using the 0.3 M-Quadrol programme, consistently gave a yield of at least 70% of that expected for the recovery of the *N*-terminal phenylthiohydantoin amino acid. Since the yield on the first cycle on each of the sequencer runs was low this made the interpretation of the sequence data, beyond cycles 36 and 26, difficult for the sequencer runs on factor \bar{D} and peptide CNBr-2 and consequently no data beyond cycles 36 and 26, in these two runs, has been reported (Table 3). If unequivocal identification of a residue was not made this has been stated in the results. The phenylthiohydantoin amino acids released were readily detected in amounts of 0.5 nmol in the sequencer runs and no problems were encountered with spurious background peaks or 'carry-over' of phenylthiohydantoin amino acids. The phenylthiohydantoin derivatives of tryptophan residues are normally readily identified by high pressure liquid chromatography and thus it is unlikely that the unidentified residue at position 22 (Figs. 4 and 5a) is tryptophan. The tryptophan at position 35 was identified by examination of the peptides produced by the selective cleavage of the tryptophanyl bonds present in factor \bar{D} (D. M. A. Johnson, J. Gagnon and K. B. M. Reid,

unpublished work). The reason why positive identification of this residue was not made from the sequencer runs reported in this paper is that position 35 is very close to the point in the runs where it was considered that unequivocal identification of the phenylthiohydantoin derivatives could not be made.

Davis *et al.* (1979a,b) have reported that the *N*-terminal amino-acid sequence of factor \bar{D} is:



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

The *N*-terminal amino-acid sequence of factor \bar{D} is shown aligned with the *N*-terminal amino acid sequences of the catalytic chains of other serine esterases in Fig. 5(a). Using the method of Moore and Goodman (1977) a preliminary statistical evaluation of the similarities between factor \bar{D} , subcomponents $\bar{C}1r$, $\bar{C}1s$, 'group-specific protease', elastase, chymotrypsin, trypsin, thrombin and factor X can be made. The method is based on the number of residues aligned and the minimum number of mutations required to make the sequences identical. Over the first 20 residues (which is all that has been published of the *N*-terminal amino acid sequences of the catalytic chains of subcomponents $\bar{C}1$ and $\bar{C}1s$) the probability of a random relationship for factor \bar{D} versus rat 'group-specific protease', elastase and subcomponent $\bar{C}1r$ or $\bar{C}1s$ was 0.01, 0.01 and 1% respectively. Comparison of subcomponent $\bar{C}1r$ with subcomponent $\bar{C}1s$ gave a value of 0.01%. The significance levels obtained on comparison (Fig. 5a) of factor \bar{D} with trypsin, chymotrypsin, plasmin, thrombin and factor X, over the first 20 residues, were all in the range $>0.1\% \leq 1\%$. Thus the *N*-terminal amino acid sequence of factor \bar{D} appears more similar to that of rat 'group-specific protease', or elastase, than to the *N*-terminal sequence of the complement subcomponents $\bar{C}1r$ and $\bar{C}1s$. This apparent strong homology of factor \bar{D} with the 'group-specific protease' from rat small intestine is of interest since it is considered that the 'group-specific protease' may have no zymogen form (Woodbury *et al.*, 1978) while it is unknown whether factor \bar{D} has a zymogen form or not. It can also be seen from Fig. 5(a) that the strictly conserved residues in the 1–36 region of the *N*-terminal sequences of the catalytic chains of other serine esterases are also conserved in factor \bar{D} , i.e. glycine at position 4, proline at 13, cystine at 26, glycine at 27 and valine at 36. The

(a)	<p>Factor D̄ 'Group-specific protease' (rat) Elastase (pig) C1r (human) 'b' chain C1s (human) 'b' chain Chymotrypsin A (cow) Plasmin (human) Trypsin (cow) Thrombin (cow) Factor X (cow)</p>	<p>1 10 20 30 36 I L G G R E A E A H A R P Y M A S V Q L - - - N X A E L C G G V L V A E Q W V I I G G V E S I P H S R P Y M A H L D I V T E K G L R V I C G G F L I S R Q F V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V I I G G Q K A K M G N F P W Q V F T N Z I I G G S D A D I K N F P W Q V F F D N I V N G E E A V P G S W P W Q V S L Q D K T G F - - - H F C G G S L I N E N W V V V G G C V A H P H S W P W Q V S L R T R F G M - - - H F C G G T L I S P E W V I V G G Y T C G A N T V P Y Q V S L - - N S G Y - - - H F C G G S L I N S Q W V I V E G Q D A E V G L S P W Q V M L F R K S P Q E - - L L C G A S L I S D R W V I V G G R D C A E G E C P W Q A L L V - N E E N E - - G F C G G T I L N E F Y V</p>
(b)	<p>Factor D̄ 'Group-specific protease' (rat) Elastase (pig) Chymotrypsin A (cow) Plasmin (human) Trypsin (cow) Thrombin (cow) Factor X (cow)</p>	<p>181 190 195 200 M C A E S - - - N R R R - D S C K G D S G G P L V C V C V G S - - P T T L R - A A F M G D S G G P L L C V C A G - - - G N G V R - S G C Q G D S G G P L H C I C A G - - - A S G V - - S S C M G D S G G P L V C L C A G H - - L A G G T - D S C Q G D S G G P L V C F C A G Y - - L E G G K - D S C Q G D S G G P V V C F C A G Y K P G E G K R G D A C E G D S G G P F V M F C A G Y - P D T Q P E - D A C Q G D S G G P H V T</p>

Fig. 5. Alignment of (a) the N-terminal region and (b) the active site of factor D̄ with N-terminal regions and active sites of other serine esterases. Sequence data for subcomponents C1r and C1s are from Sim *et al.* (1977). The sequences given for 'group-specific protease' are from Woodburg *et al.* (1978). Sequence data for chymotrypsin, trypsin, elastase, thrombin and plasmin are from de Haën *et al.* (1975) and Dayhoff (1978). a, N-Terminal sequence; the residue numbering shown is that of the factor D̄ sequence. b, Sequence around the active-site serine residue; the residue numbering shown is that of the cow chymotrypsin A sequence. —, Denotes that a gap was left to give maximum homology on alignment of the amino-acid residues is used: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, unknown; Y, Tyr.

presence of two hydrophobic residues at the *N*-terminus of the molecule has also been retained, which is consistent with the probable importance of these residues in the formation of the substrate binding site (Kraut, 1971).

In most serine proteases the active site serine is located approx. 46 amino-acid residues from the *C*-terminal end of the molecule (de Haën *et al.*, 1975). The peptide CNBr-2 is approx. 60 amino acid residues long and is located in the *C*-terminal region of the molecule (Fig. 3). This peptide was found to be labelled with radioactivity after its isolation from a CNBr digest of factor D which had been previously labelled with [1,3-¹⁴C]iPr₂P-F. The radioactive label was found to be located on the serine at position 14 of peptide CNBr-2 (Table 3) and the amino-acid sequence around this position is very typical of the highly conserved amino-acid sequence found around the active site serines of other serine esterases (Fig. 5*b*). It is probable that the relevant Met-Cys bond that was split to yield peptide CNBr-2 corresponds to the Ile₁₈₁-Cys₁₈₂ bond of bovine chymotrypsin A (Fig. 5*b*). In dogfish trypsin the equivalent portion of amino acid sequence is -Met-Met-Cys- (de Haën *et al.*, 1975).

Factor D splits factor B at a single Arg-Lys bond, thus indicating a trypsin-like specificity (Lesavre *et al.*, 1979). The presence of aspartic acid at position 8 in peptide CNBr-2 (Fig. 3) is consistent with factor D belonging to a family of trypsin-like enzymes all of which have Asp₁₈₉ (Fig. 5*b*) (chymotrypsin numbering) present in the substrate-binding pocket.

We thank Mr. T. Gascoyne for performing the amino-acid analyses and Mr. A. C. Willis and Mrs J. Parsons for excellent technical assistance. D. M. A. J. held a Commonwealth Scholarship from the Association of Commonwealth Universities.

References

- Austen, K. F. (1978) *J. Immunol.* **121**, 793-805
- Bridgen, P. J., Cross, G. A. M. & Bridgen, J. (1976) *Nature (London)* **263**, 613-614
- Davis, A. E., Zalut, C., Rosen, F. S. & Alper, C. A. (1979a) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* (Abstract) **38**, 1216
- Davis, A. E., Zalut, C., Rosen, F. S. & Alper, C. A. (1979b) *Biochemistry* **18**, 5082-5087
- Dayhoff, M. O. (ed.) (1978) *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, National Biomedical Research Foundations, Silver Springs
- de Haën, C., Neurath, H. & Teller, D. C. (1975) *J. Mol. Biol.* **92**, 225-259
- Fearon, D. T., Austen, K. F. & Ruddy, S. (1974) *J. Exp. Med.* **139**, 355-366
- Götze, O. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* (Abstract) **35**, 269
- Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124-2133
- Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126-131
- Kraut, J. (1971) *Enzymes 3rd Ed.* **3**, 165-185
- Ľaemmli, U. K. (1970) *Nature (London)* **227**, 680-685
- Lesavre, P. H. & Müller-Eberhard, H. J. (1978) *J. Exp. Med.* **148**, 1498-1509
- Lesavre, P. H., Hugli, T. E., Esser, A. F. & Müller-Eberhard, H. J. (1979) *J. Immunol.* **123**, 529-534
- Martin, A., Lachmann, P. J., Halbwachs, L. & Hobart, M. J. (1976) *Immunochemistry* **13**, 317-324
- Moore, G. W. & Goodman, M. (1977) *J. Mol. Evol.* **9**, 121-130
- Schreiber, R. D., Pangburn, M. K., Lesavre, P. H. & Müller-Eberhard, H. J. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3948-3952
- Sim, R. B., Porter, R. R., Reid, K. B. M. & Gigli, I. (1977) *Biochem. J.* **163**, 219-227
- Volanakis, J. E., Schrohenloher, R. E. & Stroud, R. M. (1977) *J. Immunol.* **119**, 337-342
- Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K. & Neurath, H. (1978) *Biochemistry* **17**, 811-819