Biochem. J. (1986) 239, 763–767 (Printed in Great Britain)

Phenotyping of human complement component C4, a class-III HLA antigen*

Edith SIM and Susan J. CROSS
Department of Pharmacology, University of Oxford, South Parks Road, Oxford OX1 3QT, and Department of Genetics, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

The plasma complement protein C4 is encoded at two highly polymorphic loci, A and B, within the class-III region of the major histocompatibility complex. At least 34 different polymorphic variants of human C4 have been identified, including non-expressed or 'null' alleles. The main method of identification of C4 polymorphic allotypes is separation on the basis of charge by agarose-gel electrophoresis of plasma. On staining by immunofixation with anti-C4 antibodies, each C4 type gives three major bands, but, since individuals can have up to five allotypes, the overlapping banding pattern is difficult to interpret. We show that digestion of plasma samples with carboxypeptidase B, which removes *C*-terminal basic amino acids, before electrophoresis, produces a single, sharp, distinct band for each allotype and allows identification of the biochemical basis of the multiple banding pattern previously observed in C4 phenotype determination.

INTRODUCTION

Human complement protein C4 is encoded at two highly polymorphic loci, C4A and C4B, within the class-III region of the major histocompatibility complex (MHC) (see Porter, 1985, for a review). At least 34 different alleles have been identified, including non-expressed, or 'null', alleles (termed 'Q0') (Mauff *et al.*, 1983), and it is known from hybridization of appropriate genomic DNA probes that some of the null alleles are associated with deletion of the C4A or C4B gene (Carroll *et al.*, 1985). Functional differences between C4A and C4B gene products have been clearly described, such that the C4B allotypes are two to three times more active in complement haemolytic assays than are C4A allotypes. C4A and C4B both contain a reactive internal thioester group, but the C4B active site is relatively more reactive with hydroxy groups (Isenman & Young, 1984; Law *et al.*, 1984; Sim & Law, 1985). This accounts for the haemolytic differences between the C4A and C4B allotypes, since hydroxy groups are plentiful on the surface of the red blood cells used as the antigen in this complement-activating system. The C4A protein is, in contrast, relatively more reactive with amino groups and is thought to play an important role in handling of immune complexes with non-cellular antigens (Law *et al.*, 1984; Isenman & Young, 1986). Clinical studies showing the association of C4AQ0 with idiopathic systemic lupus erythematosus (SLE) (Fielder *et al.*, 1983) support this conclusion.

The reactive site of C4 is contained in the C4d fragment (Campbell *et al.*, 1981), and the C4d fragment was originally shown to be polymorphic when it was identified as the component carrying the Chido(C4B) and Rodgers (C4A) blood-group antigens (Tilley *et al.*, 1978). Protein and DNA sequence studies have now defined the amino acids likely to be responsible both for the differences in functional activity and some of the antigenic differences between C4A and C4B. These are restricted to less than ten amino acids in the C4d fragment on the *C*-terminal side of the thioester site (Yu *et al.*, 1985). DNA-restriction-fragment-length polymorphisms (RFLP) have also been identified. Some correspond to known protein polymorphisms (Carroll *et al.*, 1984, 1985; Palsdottir *et al.*, 1983); others do not (Whitehead *et al.*, 1984). Since the amino acid substitutions which have been identified (14 out of 1722) (Belt *et al.*, 1984; Porter, 1985) do not correspond to known restriction-enzyme sites, it is likely that the RFLPs which have been identified are in the non-coding region. Polymorphism based on protein differences has been identified by SDS/polyacrylamide-gel electrophoresis. The α-chain of C4A is apparently 2 kDa larger than the α-chain of C4B (Roos *et al.*, 1982). This is true for all C4 allotypes which have been investigated and so does not allow distinction among allotypes of C4A or C4B. Many individuals carry a single null allele at either the A or the B locus (Awdeh *et al.*, 1979; Schendel *et al.*, 1984). Although these null heterozygotes have been detected by crossed immunoelectrophoresis (Awdeh *et al.*, 1979), which separates A allotypes from B allotypes, the method has not been used extensively, since individual C4A or C4B allotypes cannot be assigned. However, the method that has been most useful in identification of different C4 polymorphic variants has been agarose-gel electrophoresis of EDTA-plasma (see below) (Awdeh & Alper, 1980; O'Neill, 1985). The C4 pattern is identified on agarose gels after staining with anti-C4 antibodies or with a haemolytic overlay such that lysis develops where C4 is present (O'Neill, 1985). C4A gene products generally migrate more towards the anode than do C4B gene products, and this can be partly explained at the molecular level, since amino-acid-sequence studies have shown substitution of aspartate and leucine in C4A3 for histidine and arginine respectively in C4B1 (Belt *et al.*, 1984). Although neuraminidase treatment of plasma

Abbreviations used: SLE, systemic lupus erythematosus; RFLP, (DNA) restriction-fragment-length polymorphism; MHC, major histocompatibility complex.

* Dedicated to the memory of the late Professor R. R. Porter, F.R.S.

samples has improved resolution in this technique (Awdeh & Alper, 1980), one major problem in assigning patterns obtained on agarose-gel electrophoresis is that each gene product gives three bands. When several allotypes (and up to five can be inherited) are present in an individual, the bands overlap and can be difficult to interpret, even for experienced workers. This has led to controversy in assignment of rarer C4 types, for example the C4B2.9 allotype found associated with juvenile-onset diabetes (Dawkins *et al.*, 1983) and adult rheumatoid arthritis (O'Neill *et al.*, 1982). We have identified the molecular basis of this multiple-banding pattern and describe a method for converting the complex multiple-band pattern into a more informative pattern with a single sharp band for each allotype.

MATERIALS AND METHODS

Human blood was collected into EDTA-coated tubes and the separated plasma from each individual was stored at −70 °C until required. For samples of pooled EDTA-plasma, pools were made before samples were frozen. Where required, EDTA-plasma was aged, after thawing from −70 °C, by re-freezing at −40 °C for up to 2 years or by storing at 20 °C for 24 h and then for 1 week at 4 °C. Samples of plasma (50 μl) were treated either with 20 μl (6.5 units) of carboxypeptidase B (EC 3.4.17.2) in 100 mM-NaCl or with 20 μl of 100 mM-NaCl and incubated at 25 °C for 30 min. Carboxypeptidase B (Sigma, type 1) activity was determined by hydrolysis of hippurylarginine (Sigma Chemical Co., Poole, Dorset, U.K.) as described by Folk *et al.* (1960). One unit hydrolyses 1.0 μmol of hippuryl-L-arginine/min at pH 7.65 and 25 °C. Samples of plasma (18 μl) were then treated with neuraminidase (Sigma, type VIII); agarose-gel electrophoresis was carried out at pH 8.6, and C4 was detected with rabbit anti-(human C4) antibodies (Dako, High Wycombe, Bucks, U.K.), as previously described (Awdeh & Alper, 1980). Intensity of Coomassie Blue staining of C4–anti-C4 on the gel was determined by scanning gels with a Vitatron densitometer. We are grateful to Dr. M. G. Ord, Department of Biochemistry, University of Oxford, for the use of this instrument.

RESULTS

Agarose-gel electrophoresis of EDTA-plasma shows the multiple-banding pattern of C4 allotypes (Fig. 1, tracks 7–10). The overlapping bands of C4A6 and C4A3 are illustrated in track 8. However, when samples are treated with carboxypeptidase B (Fig. 1, tracks 2–5), which removes *C*-terminal arginine and lysine residues (Folk *et al.*, 1960), the pattern is simplified and a single distinct band is found for each C4 type.

Many individuals carry a single null allele at either the A or the B locus (Awdeh *et al.*, 1979; Schendel *et al.*, 1984). With the multiple-banding pattern it is difficult to assess such a hidden null allele on agarose-gel electrophoresis without carrying out family studies. However, after carboxypeptidase B digestion, identification of a hidden null allele may often be possible. In Fig. 1, track 3, the intensity of the B1 allotype is less than the intensity of any other B allotype on other tracks, and this patient, typed as carrying C4A6,A3,B1, is likely to have a hidden null allele at the B locus and the type should be C4A3,A6,B1,BQ0. Detection of heterozygotes for null alleles is further illustrated in Fig 2. Samples of EDTA-plasma from members of a family in which both

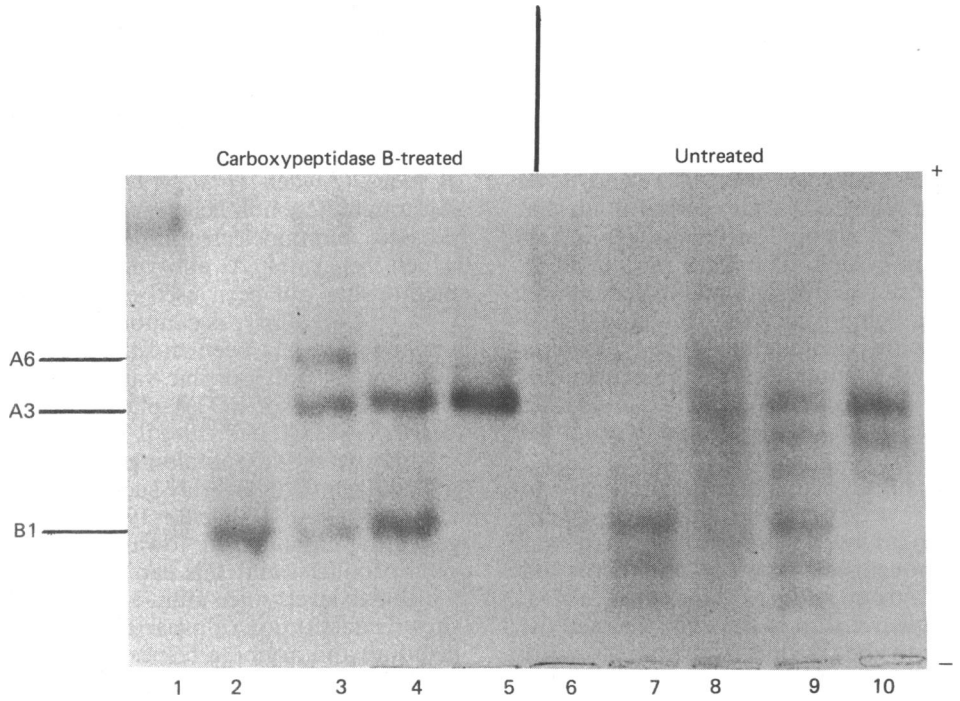

Fig. 1. Effects of carboxypeptidase B on pattern of C4 after agarose-gel electrophoresis of EDTA-plasma

Tracks 1 and 6, pooled, aged (20 °C, 4 °C) plasma; all other tracks, fresh frozen plasma for individuals. Tracks 1 and 6, C4A3,B1; 2 and 7, C4AQ0,B1; 3 and 8, C4A6,A3,B1,BQ0; 4 and 9, C4A3,B1; 5 and 10, C4A3,BQ0.

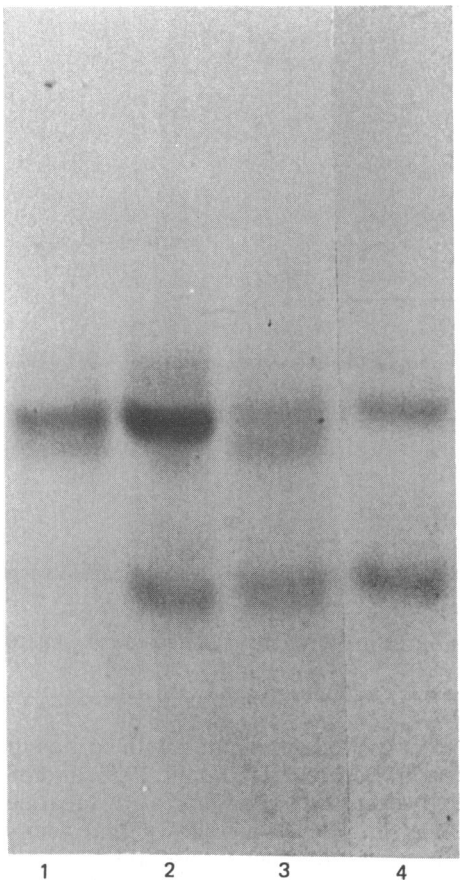

Fig. 2. Detection of heterozygotes for C4BQ0

Fresh frozen plasmas of three members of a family were treated as described in the text after carboxypeptidase B digestion. Track 1 shows the daughter with C4A3,A2,C4BQ0. Track 2 shows the mother with C4A3,C4B1,BQ0 and track 3 shows the father with C4A3,A2,C4B1,BQ0. Track 4 shows C4A3,C4B1 with no null allele for comparison.

parents carry a BQ0 allele have been treated with carboxypeptidase B before C4 typing. The father has C4A3 and C4A2 alleles and C4B1. The mother has C4A3 and C4B1. The daughter has inherited C4A3 from her mother and C4A2 from her father. The presence of a C4BQ0 allele in both parents is confirmed by the lack of any C4B gene product in their daughter. The intensity of the band at the B1 position in both parents is clearly less than the total staining intensity of the C4A bands.

Measurement of the relative intensities of staining of the A and B bands from a range of samples has shown that, in samples without a null allele, the ratio of staining intensities in the 'A region' compared with the 'B region' (A/B) is 0.9–1.2. With one null allele at the A locus, the ratio of staining intensities (A/B) is 0.5 or less. With one null allele at the B locus, A/B has been found to be between 2.6 and 3.4. These discrepancies from the 'ideal' A/B values of 0.5 or 2.0 with one null allele at the A or B locus respectively are likely to reflect lack of sensitivity in the staining and scanning procedure. However, these results do show clearly that null alleles can be readily identified without the need for family studies.

Aging of plasma samples has presented a problem in C4 typing (Fig. 1, track 6). However, treatment of an aged sample of pooled plasma with carboxypeptidase B shows up traces of bands at the A3 and B1 positions (Fig. 1, track 1).

The typing of plasma samples that have been stored at −40 °C after thawing from −70 °C is not usually reliable. However, after carboxypeptidase B treatment, the C4 types of these samples are readily determined (Fig. 3, tracks 1–3), although the intensity of the C4 bands is less than for the fresh frozen samples (Fig. 3, tracks 4–7).

Typing of samples from a range of individuals expressing the C4A3,B1 phenotype on adjacent tracks (Fig. 3, tracks 4–6) shows slight variations in the positions of these bands, although they are clearly distinct from other allotypes (Fig. 3, track 7).

DISCUSSION

In plasma, C4 consists of three polypeptide chains, α, β, and γ, linked by disulphide bridges. C4 is encoded as a single-chain precursor with the polypeptides in the order β, α, γ (Sim & Sim, 1981; Schreffler *et al.*, 1984). From analysis of the *N*-termini of these chains and complete sequence analysis of C4 cDNA, the sites of cleavage of the pro-C4 have been identified (Belt *et al.*, 1984). At these predicted cleavage sites, the *C*-terminus of the α-chain would be expected to have four *C*-terminal arginine residues and the *C*-terminus of the β-chain

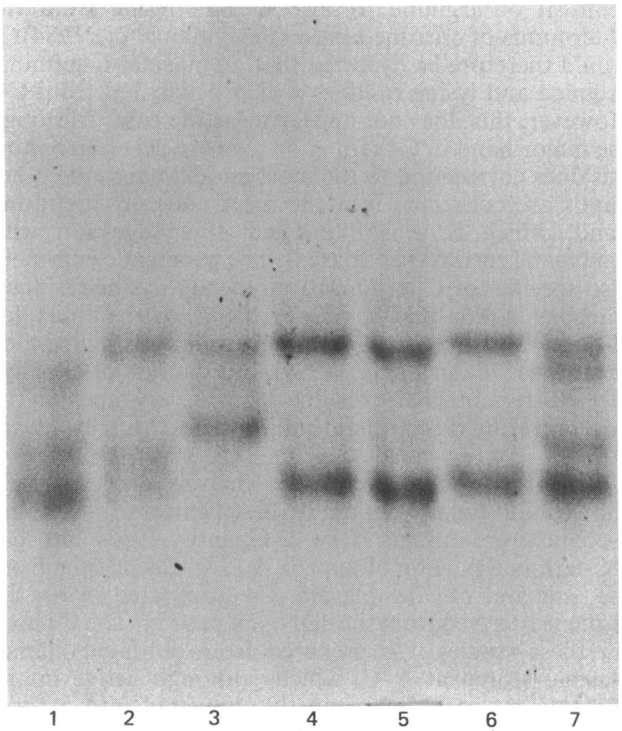

Fig. 3. Typing of C4 from samples of EDTA-plasma, thawed from −70 °C and then stored at −40 °C

Samples were treated with carboxypeptidase B before C4 typing as described in the text. Tracks 1–3, samples aged at −40 °C; tracks 4–7, fresh frozen samples. Track 1, C4A2,C4B2,B1; 2, C4A4,C4B2,B1; 3, C4A3,A1,C4BQ0; 4, C4A3,C4B1; 5, C4A3,C4B1; 6, C4A3,C4B1; 7, C4A3,A2,C4B2,B1.

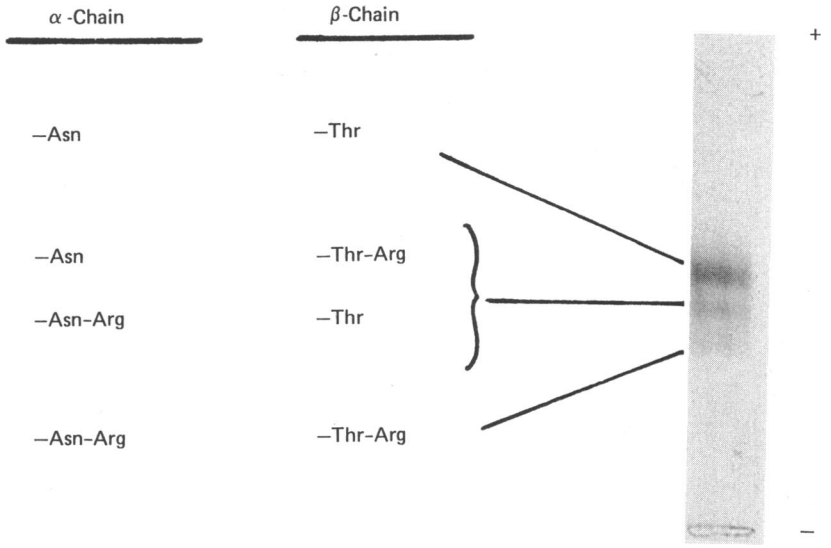

Fig. 4. Schematic diagram of the possible origin of multiple bands of C4 on agarose-gel electrophoresis without carboxypeptidase B digestion

The possible sequence at the *C*-terminus of the α- and β-chain of C4 for each of the major bands is indicated. The agarose gel shows separation of a sample of C4A3,BQ0.

would be expected to have the sequence Arg-Lys-Lys-Arg. In plasma, carboxypeptidase N activity is found (Fernandez & Hugli, 1978; McKay *et al.*, 1979; Perryman *et al.*, 1984) that is responsible for inactivation of the C5a anaphylatoxin (Fernandez & Hugli, 1978) by removal of arginine. It also cleaves lysine from the *C*-terminus of creatine kinase (Perryman *et al.*, 1984). It would therefore be expected that, in plasma, *C*-terminal arginine and lysine residues would be cleaved from C4. However, this does not appear to be the case. Although the major band in C4 typing on agarose-gel electrophoresis does correspond to the most anodal band, the other bands are converted into the most anodally migrating band, which is thus intensified after digestion with pancreatic carboxypeptidase B. The pancreatic enzyme is also specific for *C*-terminal arginine and lysine residues. Carboxypeptidase N in plasma has a lower affinity for *C*-terminal lysine than *C*-terminal arginine with dipeptide substrates (McKay *et al.*, 1979), but the rate of cleavage of the *C*-terminal basic residue depends very much on the nature of the penultimate amino acid (McKay *et al.*, 1979).

Recent *C*-terminal sequence analysis of human C4 showed that the *C*-terminal basic sequences of the β- and α-chains were missing (Law & Gagnon, 1985) and that a *C*-terminal portion of approx. 5 kDa was missing from the α-chain of the plasma C4 compared with the sequence derived from the cDNA of pro-C4. The C4 used for these studies was prepared from outdated human plasma, stored at 4 °C, which, although active in the complement cascade, is mostly converted into a form that is not informative in C4 typing by agarose-gel electrophoresis [see tracks 1 and 6 (Fig. 1) in which plasma has been aged artificially]. The three bands found on agarose-gel electrophoresis without carboxypeptidase B digestion may represent the cleavages illustrated in Fig. 4. Although other cleavages are possible, it seems likely that the difference in migration of each of the bands in agarose-gel electrophoresis without carboxypeptidase B digestion represents a single charge difference. Electrophoresis of native C5a compared with carboxypeptidase B-treated C5a (Fernandez & Hugli, 1978) and of creatine kinase and carboxypeptidase-treated creatine kinase (Perryman *et al.*, 1984) has shown that removal of a single charged residue gives rise to a mobility shift similar to that reported here.

Many disease associations with MHC antigens have been described, and it has been proposed that the C4 type in linkage disequilibrium with particular MHC haplotypes, e.g. HLA A1 B8 Dr3 C4AQ0B1 in association with SLE (Fielder *et al.*, 1983), may be the informative antigen in the disease association (Porter, 1983). The value of matching of organ donors and recipients across the MHC rather than at only the class II (Dr) region has been suggested on very strong evidence showing the lack of a mixed lymphocyte reaction in such unrelated matched pairs (Awdeh *et al.*, 1985). Since C4 is the most polymorphic class-III antigen, C4 typing would be very useful in this respect and also in assessing the importance of C4 in MHC disease associations.

There are at present less than 15 centres in the world which routinely carry out C4 typing (Mauff *et al.*, 1983). This is undoubtedly because of problems in interpretation of C4 typing patterns and the requirement for family studies for reliable types to be obtained. This is especially true for samples that have not been frozen rapidly. The method we describe is helpful in all of these respects and we hope it will encourage more laboratories to carry out C4 typing.

We thank Dr. R. B. Sim for helpful discussions, Professor J. H. Edwards, F.R.S., and the Medical Research Council for providing laboratory facilities for C4 typing, Miss M. Wood for technical assistance and the Wellcome Trust for financial help.

REFERENCES

Awdeh, Z. L. & Alper, C. A. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 3576–3580

Awdeh, Z. L., Raum, D. & Alper, C. A. (1979) Nature (London) **282**, 205–207

Awdeh, Z. L., Alper, C. A., Eynon, E., Alosco, S. M., Stein, R. & Yunis, E. J. (1985) Lancet **ii**, 853–855
Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) Cell (Cambridge, Mass.) **36**, 907–914
Campbell, R. D., Gagnon, J. & Porter, R. R. (1981) Biochem. J. **199**, 359–370
Carroll, M. C., Belt, T., Palsdottir, A. & Porter, R. R. (1984) Philos. Trans. R. Soc. London Ser. B **306**, 379–388
Carroll, M. C., Palsdottir, A., Belt, K. T. & Porter, R. R. (1985) EMBO J. **4**, 2547–2552
Dawkins, R. L., Christiansen, F. T., Kay, P. H., Garlepp, M., McCluskey, J., Hollingsworth, P. N. & Zilko, P. J. (1983) Immunol. Rev. **70**, 5–22
Fernandez, H. N. & Hugli, T. E. (1978) J. Biol. Chem. **253**, 6955–6962
Fielder, A. H. L., Walport, M. J., Batchelor, J. R., Rynes, R. I., Black, C. M., Dodi, I. A. & Hughes, G. R. V. (1983) Br. Med. J. **286**, 425–428
Folk, J. E., Piez, K. A., Carroll, W. R. & Glander, J. A. (1960) J. Biol. Chem. **235**, 2272–2277
Isenman, D. E. & Young, J. R. (1984) J. Immunol. **132**, 3019–3027
Isenman, D. E. & Young, J. R. (1986) J. Immunol. **136**, 2542–2550
Law, S. K. A. & Gagnon, J. (1985) Biosci. Rep. **5**, 913–921
Law, S. K. A., Dodds, A. W. & Porter, R. R. (1984) EMBO J. **3**, 1819–1823
Mauff, G., Alper, C. A., Awdeh, Z., Batchelor, J. R., Bertrams, T., Braun-Pettersen, G., Dawkins, R. L., Demant, P., Edwards, J., Grosse-Wilde, H., Hauptmann, G., Klonda, P., Lamm, L., Mollenhauer, E., Nerl, C., Olaisen, B., O'Neill, G., Rittner, C., Roos, M. H., Skanes, V., Teisberg, P. & Wells, L. (1983) Immunobiology **164**, 184–191
McKay, T. J., Phelan, A. W. & Plummer, T. H. (1979) Arch. Biochem. Biophys. **197**, 487–492
O'Neill, G. J. (1985) in Methods in Complement for Clinical Immunologists (Whaley, K., ed.), pp. 282–283, Churchill–Livingstone, Edinburgh
O'Neill, G. J., Nerl, C. W., Kay, P. H. Christiansen, F. T., McCluskey, J. & Dawkins, R. L. (1982) Lancet **ii**, 214–215
Palsdottir, A., Cross, S. J., Edwards, J. H. & Carroll, M. C. (1983) Nature (London) **306**, 615–616
Perryman, M. B., Knell, J. D. & Roberts, R. (1984) Clin. Chem. **30**, 662–664
Porter, R. R. (1983) Mol. Biol. Med. **1**, 161–168
Porter, R. R. (1985) Immunol. Rev. **87**, 7–17
Roos, M. H., Mollenhauer, E., Demant, P. & Rittner, C. (1982) Nature (London) **298**, 854–856
Schendel, D. J., O'Neill, G. J. & Wan, K. R. (1984) Immunogenetics **20**, 23–31
Schreffler, D. C., Atkinson, J. P., Chan, A. C., Karp, D. R., Killion, C. C., Ogata, R. T. & Rosa, P. A. (1984) Philos. Trans. R. Soc. London Ser. B **306**, 395–403
Sim, E. & Law, S. K. A. (1985) FEBS Lett. **184**, 323–327
Sim, R. B. & Sim, E. (1981) Biochem. J. **193**, 129–141
Tilley, C. A., Romans, D. G. & Crookston, M. C. (1978) (London) **276**, 713–715
Whitehead, A. S., Woods, D. E., Fleischnick, E., Chin, J. E., Yunis, E. J., Katz, A. J., Gerald, P. S., Alper, C. A. & Colten, H. R. (1984) N. Engl. J. Med. **310**, 88–91
Yu, C.-Y., Belt, K. T., Giles, C. M. & Porter, R. R. (1985) Complement **2**, 89–90

Received 22 July 1986; accepted 19 August 1986