Biochem. J. (1991) 277, 387–391 (Printed in Great Britain)

Complete nucleotide and deduced amino acid sequence of human β_2-glycoprotein I

Alexander STEINKASSERER,*‡ Cornelia ESTALLER,† Elisabeth H. WEISS,† Robert B. SIM* and Anthony J. DAY*

*M.R.C. Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and †Institute of Immunology, University of Munich, Goethestrasse 31, 8000 Munich-2, Federal Republic of Germany

The nucleotide and complete amino acid sequence for the human β_2-glycoprotein I (β_2I) was derived by sequencing the cDNA clone pB$_2$I-1. In addition to the 326 amino acid residues of the mature protein this clone codes for a putative leader peptide and contains sequence representing 5′ and 3′ untranslated regions. When this amino acid sequence was compared with the previously published primary sequence, three major amino acid substitutions were found, two involving cysteine residues. These substitutions lead to a new alignment of the complement control protein (CCP) repeats present in β_2I and a prediction of the complete disulphide bond organization. Northern-blot analysis indicates that hepatocytes are a major site of biosynthesis for this protein. A transcription signal of about 1.5 kb was detected by using RNA from HepG2 cells.

INTRODUCTION

β_2-Glycoprotein I (β_2I) is a protein of approx. 50 kDa that was first described by Schultze *et al.* (1961). It is associated with lipoproteins in plasma at a concentration of about 200 μg/ml (Polz & Kostner, 1979). The site of biosynthesis is not known. The amino acid sequence of β_2I was originally determined by protein-sequencing techniques and is 326 residues in length (Lozier *et al.*, 1984). The sequence consists of five contiguous repeating units, of about 60 amino acid residues, that belong to the complement control protein (CCP) superfamily (Day *et al.*, 1989; Reid & Day, 1989). The CCP repeat is present in the proteins involved in the regulation of complement activation and in a growing number of non-complement proteins, including interleukin-2 receptor and a new family of leucocyte adhesion molecules (Day *et al.*, 1989).

Although the function of β_2I has been investigated, no clear physiological role has yet been determined. However, β_2I has been shown to bind to anionic phospholipids (Wurm, 1984), platelets (Schousboe, 1980), heparin (Polz *et al.*, 1979), DNA (Kroll *et al.*, 1976) and mitochondria (Schousboe, 1979). A possible role of β_2I is inhibition of the intrinsic blood coagulation cascade by binding to negatively charged phospholipids on the surface of damaged cells, thus preventing activation of the pathway (Schousboe, 1985; Nimpf *et al.*, 1986). Furthermore, it inhibits both prothrombinase activity of activated platelets (Nimpf *et al.*, 1986) and ADP-mediated platelet aggregation (Nimpf *et al.*, 1987). McNeil *et al.* (1990) and Galli *et al.* (1990) have demonstrated that anti-phospholipid antibodies, present in certain autoimmune diseases (e.g. systemic lupus erythematosus), are directed against an antigen consisting of a complex of β_2I and negatively charged phospholipids (cardiolipin).

As no structural information at the DNA level was available, we have determined the complete cDNA sequence for β_2I and provide evidence for a site of biosynthesis.

EXPERIMENTAL

Polymerase chain reaction

Total RNA was isolated from an HepG2 human liver cell line by using guanidinium thiocyanate extraction and a CsCl gradient (Maniatis *et al.*, 1982). A 10 μg portion was used for the first-strand cDNA synthesis (cDNA synthesis kit; Amersham, Aylesbury, Bucks., U.K.). One-tenth of this reaction mixture was used for PCR amplification. The sequences of the synthetic oligonucleotides used to amplify the β_2I transcript were deduced from the primary amino acid sequence (Lozier *et al.*, 1984) according to the human codon bias described by Lathe (1985). The following oligonucleotides were generated by using an Applied Biosystems 381A DNA synthesizer: sense 5′-GGCCGGACCTGCCCCAAGCCTGATGACCTGCCATT and antisense 5′-GCATGGCTTCACATCGGAGGCATC-TGTCTTCCA. Twenty-five cycles of amplification (1 min at 95 °C, 1 min at 60 °C and 3 min at 72 °C) were performed in a 100 μl reaction mixture [10 mM-Tris/HCl buffer, pH 8.3, containing 50 mM-KCl, 1.5 mM-MgCl$_2$, 0.01% gelatin, 200 μM-dNTP, 1 μM-5′-primer, 1 μM-3′-primer and 3 units of *Thermus aquaticus* (*Taq*) polymerase (Perkin–Elmer Cetus)] covered with 100 μl of paraffin. A 5 μl portion of this reaction mixture was analysed by electrophoresis on a 0.7% (w/v) agarose gel.

Isolation of β_2I clones and cDNA sequencing

A human liver cDNA library constructed in the eukaryotic expression vector CDM8 (Estaller *et al.*, 1991) was screened with the β_2I PCR product (978 bp in length). Approx. 300000 colonies were screened with the PCR probe, which had been labelled with [α-^{32}P]ATP by the polypriming method (Feinberg & Vogelstein, 1983). The length of the cDNA inserts from positive clones were estimated by a *Pst*I–*Hin*dIII restriction digest. The sequence of a 1.15 kb insert (designated clone pB$_2$I-1) was determined from both strands by using a double-strand ^{T7}Sequencing Kit (Pharmacia, Milton Keynes, Bucks., U.K.). Segments of a further six positive clones were also sequenced.

Northern-blot analysis

A 10 μg portion of total RNA was electrophoresed in a 1.2% (w/v) agarose gel in Mops buffer containing 5% (v/v) formaldehyde (Maniatis *et al.*, 1982). RNA was transferred to a Hybond-N membrane (Amersham) and hybridized at 65 °C with

Abbreviations used: β_2I, β_2-glycoprotein I; CCP, complement control protein.

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X58100.

the [α-^{32}P]ATP-radiolabelled pB_2I-1 probe. The blots were washed at high stringency as described by Church & Gilbert (1984).

Computer analysis

The five CCP repeats of β_2I (CCP repeat 1, residues 1–61; CCP repeat 2, residues 62–119; CCP repeat 3, residues 120–182; CCP repeat 4, residues 183–242; CCP repeat 5, residues 243–307) were aligned by the computer program Alignment of Multiple Protein Sequences (AMPS) (Barton & Sternberg, 1987) with the mutation data matrix of Dayhoff *et al.* (1983) with a bias of six added to each term of the matrix and a break penalty of six.

RESULTS

PCR was used to generate a β_2I transcript from HepG2 RNA. A single 978 bp product was obtained by using synthetic oligonucleotide primers and this was used to screen a human liver cDNA library that was subcloned into the CDM8 expression vector. Several hundred colonies hybridized specifically with the 978 bp PCR probe. Ten colonies were picked and were positive on re-screening. All contained an approx. 1.1 kb-long *Pst*I–*Hin*dIII insert that hybridized specifically with the 978 bp probe. The cDNA insert of one clone (pB_2I-1) was sequenced by plasmid double-strand sequencing with the use of two oligonucleotides located 5′ and 3′ to the CDM8 subcloning site (Estaller *et al.*, 1991). Additional synthetic oligonucleotides based on internal sequences were generated to determine the nucleotide sequence from both strands. The sequencing strategy used is shown in Fig. 1.

The complete nucleotide and deduced amino acid sequence of clone pB_2I-1 is illustrated in Fig. 2. The clone pB_2I-1 encodes the complete protein sequence of the mature human β_2I. Furthermore, the cDNA insert contains 21 bp of the 5′ untranslated region followed by coding sequence for a 19-amino acid-residue putative leader peptide starting with an ATG triplet. The coding sequence for the mature protein (326 amino acid residues) is 978 bp long (beginning at base 79) and is followed by a TAA stop codon. The 93 bp-long 3′ untranslated region contains a putative polyadenylation signal (AATAAA) 73 bp downstream from the stop codon (see Fig. 2). Comparing the published protein sequence (Lozier *et al.*, 1984) with the deduced amino acid sequence from clone pB_2I-1, three major differences were found. The cysteine at amino acid position 102 is substituted in the cDNA-derived sequence by a serine residue, asparagine at position 169 is replaced by a cysteine and the valine at position 247 is changed to a leucine residue, as indicated on Fig. 2. Sequencing of the appropriate regions of six further clones confirmed that in each case a serine was encoded at amino acid position 102, and a cysteine at position 169. At position 247, however, leucine (TTA) was found in six clones (including pB_2I-1) and valine (GTA) was found in one clone (pB_2I-2).

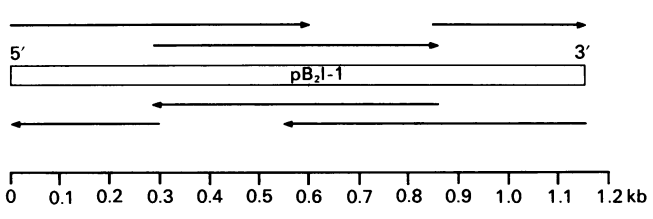

Fig. 1. Sequencing strategy

The cDNA insert of clone pB_2I-1 was sequenced by plasmid double-strand sequencing using 5′- and 3′-oligonucleotides located in the CDM8 subcloning site. Internal oligonucleotides were synthesized for multiple sequencing from both strands. Arrows indicate the extent and direction of DNA sequence determination.

The site of synthesis of β_2I was investigated by Northern-blot analysis. Total RNA was isolated from the cell lines HepG2, Hela, Raji and Molt4 and also from U937 cells, which were cultured for 3 days with and without phorbol 12-myristate 13-acetate (30 ng/ml). A 10 μg portion of each RNA was separated on a 1.2% (w/v) agarose gel, which was subsequently blotted and hybridized with the 1.15 kb pB_2I-1 probe. A signal corresponding to an approx. 1.5 kb transcript was detected in HepG2-derived RNA, whereas no signal was detected in the RNA from the other cell lines (results not shown). The strong hybridization signal after a short exposure (5 h) indicated a large amount of β_2I mRNA present in HepG2 cells. Even after long exposure (2 days) it was not possible to detect any signal in the other RNAs tested. The abundance of positive clones in the liver cDNA library and the presence of β_2I mRNA in the hepatoma cell line indicate that hepatocytes are a major site of β_2I synthesis.

DISCUSSION

The complete nucleotide sequence and derived amino acid sequence for human β_2I has been established by cDNA cloning and sequencing. There are two major differences, both involving cysteine residues, between the cDNA-derived amino acid sequence of β_2I and that established by protein-sequencing techniques (Lozier *et al.*, 1984). The cysteine at position 102 is replaced by a serine residue and asparagine-169 is replaced by a cysteine residue in the derived amino acid sequence. The two changes involving cysteines occur in CCP repeats 2 and 3 of β_2I. The CCP repeat is based on a consensus sequence with four invariant cysteine residues disulphide-bonded in a pattern Cys 1–3 and Cys 2–4 (Day *et al.*, 1989; Janatova *et al.*, 1989; Day, 1991). In the protein sequence published by Lozier *et al.* (1984), CCP repeat 2 had a cysteine residue additional to the consensus cysteine residues and CCP repeat 3 had consensus Cys 3 missing. Therefore CCP repeats 2 and 3 of β_2I were unusual examples of CCP repeats in general. It is unlikely that the differences involving cysteine residues are polymorphisms, as these are structurally important residues, and were probably incorrect assignments in the protein sequence. However, the third difference noted, valine to leucine at position 247, is likely to be a polymorphism: both valine and leucine were found to be encoded in different clones from the same cDNA library, suggesting that the DNA donor was heterozygous.

A partial cDNA sequence for rat β_2I has been published in abbreviated format (Aoyama *et al.*, 1989). In comparison with the sequence shown in Fig. 2, a large segment, corresponding to bases 88–232 of Fig. 2, appears to be missing from the rat clone. The first 73 bases of the rat sequence align with nucleotide residues 15–87 of Fig. 2, with 13 mismatches. This region corresponds to the last seven nucleotide residues of the 5′ untranslated region, the entire signal sequence and the first two amino acid residues of the mature protein. The rest of the rat sequence aligns with nucleotide residues 232–1136 in Fig. 2, with 182 mismatches. Aoyama *et al.* (1989) suggested that rat β_2I has major differences from the human protein at the *N*-terminal end, but this appears likely to have arisen from a cloning artifact involving an approx. 145 bp deletion. The deletion does not correspond to a complete domain at the protein level. The major portion of the rat cDNA-derived amino acid sequence corresponds to amino acid residues 52–326 of the human sequence (Fig. 2). The rat and human sequences are 82.5% identical over this region, with all cysteine residues, including position 169, conserved. Position 102 is also a serine residue in

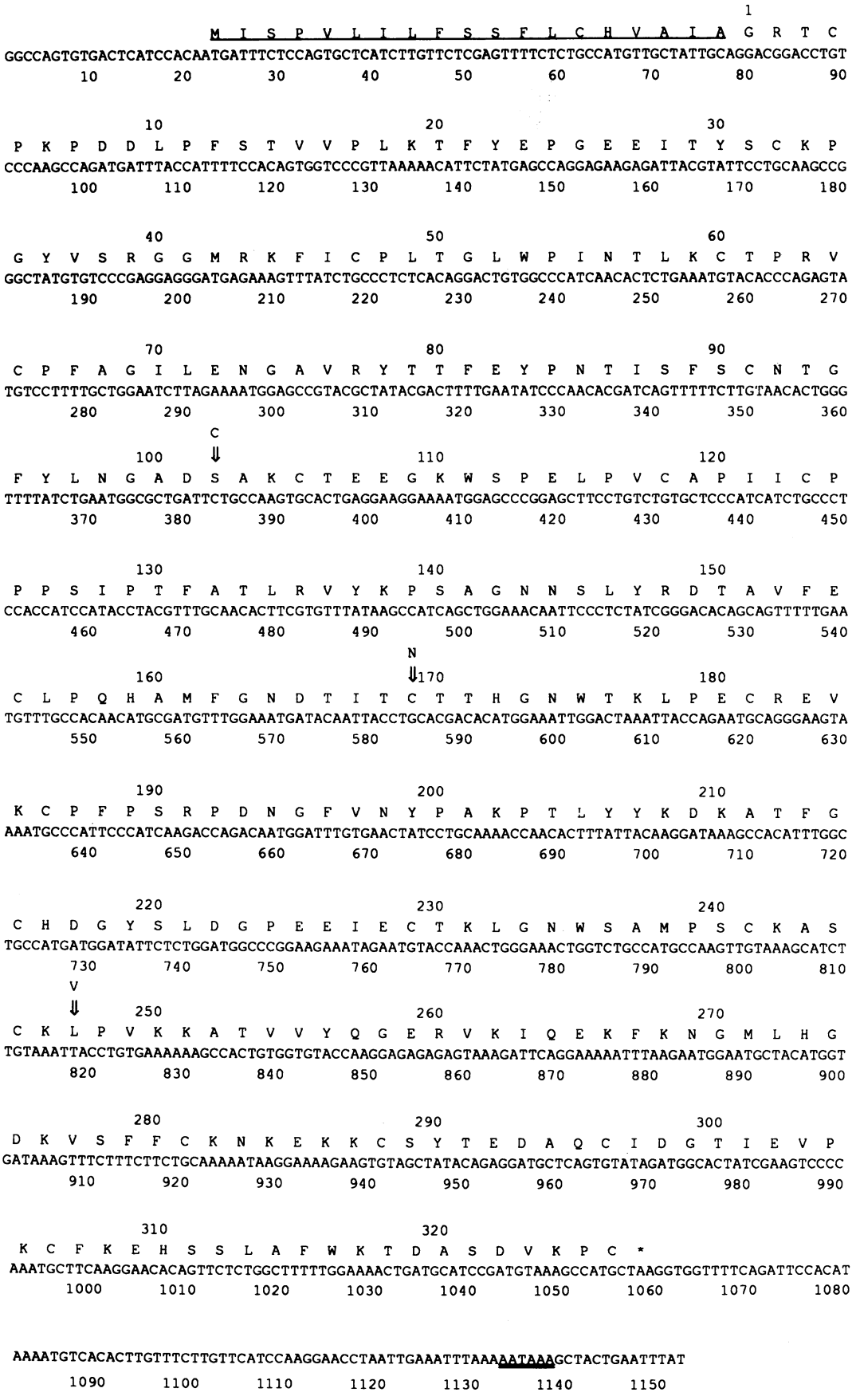

Fig. 2. Nucleotide sequence and deduced amino acid sequence of β_2I

The cDNA-derived amino acid sequence is shown. The putative 19-amino acid-residue leader sequence starting with a methionine residue is underlined. Vertical arrows indicate the three amino acid differences at positions 102 (Cys→Ser), 169 (Asn→Cys) and 247 (Val→Leu) when compared with the protein sequence published by Lozier *et al.* (1984). The possible polyadenylation signal is double-underlined.

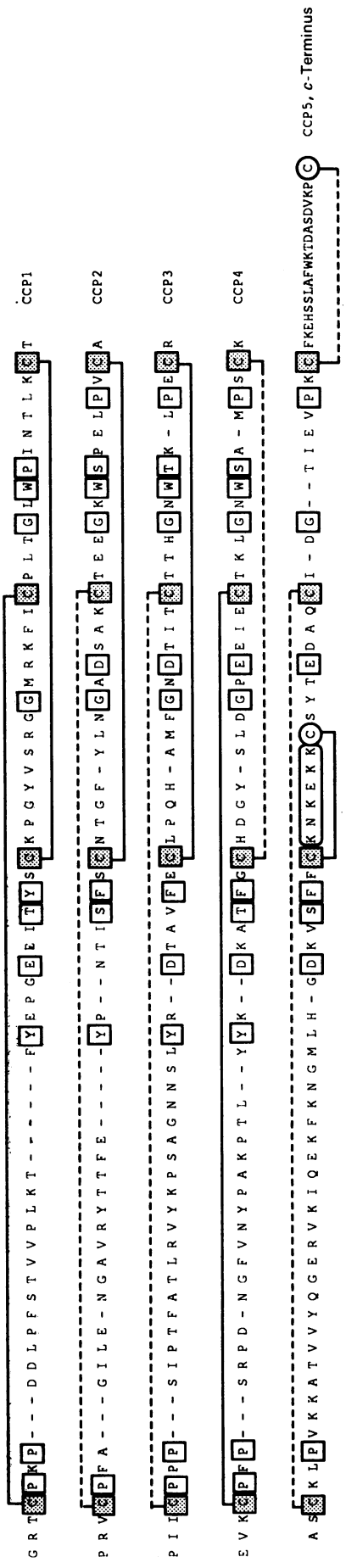

Fig. 3. Alignment and disulphide bond organization of β_2I

The five CCP repeats of β_2I were aligned using the computer program AMPS (Barton & Sternberg, 1987). The four consensus cysteine residues of each CCP repeat (generally disulphide-bonded in a pattern Cys 1–3 and Cys 2–4) are indicated by shaded boxes. Additional non-consensus cysteines are circled. Disulphide bonds mapped by Lozier *et al.* (1984) are indicated by continuous lines and further predicted disulphide bonds are shown by broken lines. Other identities and conservative replacements (present in at least four out of five repeats) are indicated by open squares. The highly positively charged region (KNKEKK) present in CCP repeat 5 is denoted by a box and might be the interaction site for negatively charged phospholipids.

the rat sequence, and the 'polymorphic' position 247 is a leucine residue.

The cDNA-deduced amino acid sequence establishes CCP repeats 2 and 3 of β_2I as being typical members of the CCP repeat superfamily, each having four cysteine residues. A new alignment of the repeats in β_2I is shown in Fig. 3. CCP repeats 2 and 3 have one disulphide bridge mapped (Cys 2–4 in both cases), and it is very likely that they will have the normal Cys 1–3 linkages (see Fig. 3). In contrast with the first four CCP repeats of β_2I, which are typical examples of this domain, the fifth repeat is the most aberrant member of the CCP superfamily. Nevertheless, it fulfils the criteria for membership of this family (Patthy, 1987). In CCP repeat 5 there is an extra non-consensus cysteine residue, disulphide-linked to consensus Cys 2 (see Fig. 3). It seems likely that the consensus Cys 1–3 disulphide linkage will form normally and that consensus Cys 4 is linked to the cysteine residue at the *C*-terminal end of the protein (see Fig. 3). A CCP repeat from complement Factor H has been shown to be an independently folding domain and its secondary structure has been determined by two-dimensional n.m.r. spectroscopy (Barlow *et al.*, 1991). Although CCP repeat 5 has a unique disulphide bond organization and an additional *C*-terminal region, it is likely that these differences can be accommodated within the secondary-structure model of the CCP domain.

β_2I in association with negatively charged phospholipids has been identified as an autoantigen involved in certain autoimmune disorders (e.g. systemic lupus erythematosus) (McNeil *et al.*, 1990; Galli *et al.*, 1990). In this regard CCP repeat 5 of human and rat β_2I contains a highly positively charged sequence KNKEKK (residues 282–287) that is constrained by a disulphide bridge (see Fig. 3). On the basis of the secondary structure model for the CCP domain (Barlow *et al.*, 1991), this motif is likely to be present as a surface-exposed turn. Therefore it is a possibility that this sequence may constitute the binding site for negatively charged phospholipids.

In the original sequence paper by Lozier *et al.* (1984), five potential asparagine-linked glycosylation sites were proposed. As asparagine-169 is replaced by a cysteine residue in the cDNA-derived protein sequence, this can no longer constitute a sugar-attachment site. Work by Walsh *et al.* (1990) suggests that the potential carbohydrate-attachment sites are all utilized and the sugars are of bi- and tri-antennary type.

The cloning and sequencing of β_2I allows the use of protein-engineering techniques to investigate its physiological function. There is a vital need to understand the involvement of β_2I in autoimmune disorders such as systemic lupus erythematosus. In this regard, further work is required to express the fifth CCP repeat, together with the *C*-terminal portion, to assess whether the proposed phospholipid interaction site of β_2I is present in this region.

We thank Dr. S. A. Krilis for generously providing information before publication. We also gratefully acknowledge B. Moffatt for providing cell cultures, Dr. M. Baron and Dr. C. Milner for criticism of the manuscript and Dr. D. Norman for helpful discussion. This work is a contribution from the Oxford Centre for Molecular Sciences (O.C.M.S.), which is supported by the Science and Engineering Research Council and the Medical Research Council. A.S. was supported by an O.C.M.S. post-doctoral fellowship.

REFERENCES

Aoyama, Y., Chan, Y. L. & Wool, I. G. (1989) Nucleic Acids Res. **17**, 6401

Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B. & Campbell, I. D. (1991) Biochemistry **30**, 997–1004

Barton, G. J. & Sternberg, M. J. (1987) J. Mol. Biol. **198**, 327–337

Church, G. M. & Gilbert, W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1991–1995
Day, A. J. (1991) in Biochemistry and Molecular Biology of Complement (Sim, R. B., ed.), Kluwer Academic Press, Lancaster, in the press
Day, A. J., Campbell, R. D. & Reid, K. B. M. (1989) in Progress in Immunology VII (Melchers, F., ed.), pp. 209–212, Springer-Verlag, Heidelberg
Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Methods Enzymol. **91**, 524–545
Estaller, C., Schwaeble, W., Dierich, M. P. & Weiss, E. H. (1991) Eur. J. Immunol. **21**, 799–802
Feinberg, A. P. & Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13
Galli, M., Comfurius, P., Maassen, C., Hemker, H. C., De Baets, M. H., van Breda-Vriesman, P. J. C., Barudi, T., Zwaal, R. F. A. & Breves, E. M. (1990) Lancet **i**, 1544–1547
Janatova, J., Reid, K. B. M. & Willis, A. C. (1989) Biochemistry **28**, 4754–4761
Kroll, J., Larsen, J. K., Loft, H., Ezban, M., Wallevik, K. & Faber, M. (1976) Biochim. Biophys. Acta **434**, 490–501
Lathe, R. (1985) J. Mol. Biol. **183**, 1–12
Lozier, J., Takahashi, N. & Putnam, F. W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 3640–3644
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor
McNeil, P. H., Simpson, R. J., Chesterman, C. H. & Krilis, S. A. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 4120–4124
Nimpf, J., Bevers, E. M., Bomans, P. H. H., Till, U., Wurm, H., Kostner, G. M. & Zwaal, R. F. A. (1986) Biochim. Biophys. Acta **884**, 142–149
Nimpf, J., Wurm, H. & Kostner, G. M. (1987) Atherosclerosis **63**, 109–114
Patthy, L. (1987) J. Mol. Biol. **198**, 567–577
Polz, E. & Kostner, G. M. (1979) FEBS Lett. **102**, 183–186
Polz, E., Wurm, H. & Kostner, G. M. (1979) Int. J. Biochem. **11**, 265–273
Reid, K. B. M. & Day, A. J. (1989) Immunol. Today **10**, 177–180
Schousboe, I. (1979) Biochim. Biophys. Acta **579**, 396–408
Schousboe, I. (1980) Thromb. Res. **19**, 225–237
Schousboe, I. (1985) Blood **66**, 1086–1091
Schultze, H. E., Heide, H. & Haupt, H. (1961) Naturwissenschaften **48**, 719–724
Walsh, M. T., Watzlawick, H., Putnam, F. W., Schmid, K. & Brossmer, R. (1990) Biochemistry **29**, 6250–6257
Wurm, H. (1984) Int. J. Biochem. **16**, 511–515

Received 3 October 1990/14 January 1991; accepted 5 February 1991