Volume 313, number 2, 193–197 FEBS 11780 November 1992

Activity, disulphide mapping and structural modelling of the fifth domain of human β_2-glycoprotein I

Alexander Steinkasserer[a], Paul N. Barlow[b], Anthony C. Willis[a], Zsuzsa Kertesz[a], Iain D. Campbell[b], Robert B. Sim[a] and David G. Norman[c]

[a]*MRC Immunochemistry Unit, Department of Biochemistry, and* [b]*Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK* and [c]*Department of Biochemistry, Medical Science Institute, University of Dundee, Dundee, DD1 4HN, UK*

Received 1 October 1992; revised version received 8 October 1992

Complexes formed by the interaction of negatively charged phospholipids and β_2-glycoprotein I (β_2-I) are the target of autoantibodies in systemic lupus erythematosus. The highly positively charged fifth (C-terminal) domain of human β_2-I was produced as a fusion protein in an *Escherichia coli* expression system and was shown to bind to the negatively charged phospholipid, cardiolipin, almost as well as the intact protein. In an attempt to define the 3D structure of this domain, the disulphide linkage pattern was determined and shown to be Cys 1-4, Cys 2-5 and Cys 3-6 in contradiction to an earlier report. In the light of this information, the sequence of the fifth domain of β_2I (β_2-I-5) is readily aligned with that of the 16th repeat of factor H, of which the 3D structure is known, and a model of β_2I-5 has been built by homology. On the basis of the model we suggest residues that might be the target of profitable site-directed mutagenesis in structure–function studies

β_2-Glycoprotein I (β_2-I); Apolipoprotein H (APOH); Cardiolipin; Disulphide linkage; Homology modelling; Complement control protein (CCP)

1. INTRODUCTION

β_2-Glycoprotein I (β_2-I) is associated with different fractions of lipoproteins in the plasma [1]. This protein, also known as apolipoprotein H (APOH), was first described by Schultze et al. [2] and the amino acid sequence was determined by Lozier et al. [3]. The complete nucleotide sequence has been established by cDNA sequencing [4], and using somatic cell hybrids the β_2-I gene was localized on chromosome 17q23-25 [5].

The physiological function of β_2-I is not fully established, but it has been shown to bind to DNA [6], mitochondria [7], platelets [8] and to negatively charged phospholipids [9]. Recently it has been shown that a complex of β_2-I and negatively charged phospholipids forms an autoantigen recognized by certain antibodies present in autoimmune diseases like systemic lupus erythematosus (SLE) [10,11].

The β_2-I protein is composed of five repeating domains of some 60–80 amino acids residues each. Within each repeat it is possible to recognise the consensus sequence of the complement control protein (CCP) module [4], implying that β_2-I is a member of the CCP superfamily. CCP modules are found in many complement proteins, including factor H, CR1, CR2, C4bp, as well as in some non-complement proteins such as the IL-2 receptor and the selectin family of endothelial leukocyte adhesion molecules [12]. The three-dimensional solution structures of three CCP modules have now been established by NMR ([13,14] and unpublished data) and are seen to be very similar, confirming that a consensus sequence confers extensive topological similarities. This observation means that it is feasible to create accurate models, by homology, for other members of the CCP superfamily without recourse to NMR spectroscopy or X-ray crystallography methods.

In this study we focus on the fifth β_2-I domain (β_2I-5). This module contains 15 lysine residues and, since β_2-I binds to negatively charged phospholipids (e.g. cardiolipin), this β_2I-5 domain may be important in the protein–lipid interaction. To address this hypothesis, β_2I-5 was expressed as a recombinant fusion protein in *Escherichia coli*, and has been shown to bind to cardiolipin. In addition we present the correct disulphide mapping of the six cysteines present in this repeat, originally incorrectly reported [3]. Finally we report a model for the tertiary structure of β_2I-5 built by homology with the 16th module of factor H.

2. MATERIALS AND METHODS

2.1 *Recombinant expression and binding studies of β_2I-5 fusion protein to cardiolipin*

The coding region for the fifth domain (amino acid residues 243–326) was amplified by PCR from the cDNA clone pβ_2I-I [4], whit the

Correspondence address: A. Steinkasserer, MRC Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford, OX1 3QU, UK. Fax: (44) (865) 275 729.

substitution of a glycine residue for Ala (243), the introduction of a *Bam*HI restriction site at the 5' end, and an *Eco*RI site at the 3' end. The PCR product was subcloned into the fusion vector pGEX-2T [15] and transformed into the *E. coli* strain, NM554. An overnight bacterial culture was diluted 1:10 in fresh LB medium and after 1 h the expression of the fusion protein (glutathione *S*-transferase-β_2I-5 or GST-β_2I-5) was induced with 0.1 mM IPTG and the culture grown for a further 3 h. The cells were pelleted, lysed and incubated with glutathione agarose beads in order to bind the fusion protein onto the beads [16]. The recombinant GST-β_2I-5 was eluted from the beads by competition with free glutathione (5 mM). After spinning down the beads the fusion protein was recovered in the supernatant and used for binding studies to cardiolipin.

A competitive inhibition assay of the binding of ^{125}I-labelled β_2-I to cardiolipin by unlabelled β_2-I and by GST-β_2I-5 was performed. The conditions for the assay were as described previously [9]. Briefly, cardiolipin (50 μg/ml in ethanol) was coated onto microtiter plates (100 μl/well) and the ethanol allowed to evaporate. Non-specific binding sites were blocked with 250 μg/ml α_2-macroglobulin (200 μl/well). Human β_2-I was isolated as previously described [9] and radiolabelled with Na^{125}I using iodobeads. Labelled β_2-I (50 μl; 0.88 μg; 1.35×10^6 cpm) in 10 mM potassium phosphate, 1.5 mM $CaCl_2$, 0.5 mM $MgCl_2$, pH 7.0, containing 20 μg/ml bovine serum albumin was preincubated (30 min, 20°C) with 50 μl of the same buffer containing various amounts of unlabelled β_2-I or unlabelled GST-β_2I-5. These mixtures were then transferred to cardiolipin-coated wells, and incubated for 30 min at 20°C. Wells were then washed with 4 × 200 μl of 10 mM potassium phosphate, 1.5 mM $CaCl_2$, 0.5 mM $MgCl_2$, pH 7.0, containing 20 μg/ml bovine serum albumin, and bound radioactivity was eluted with 200 μl of 0.1 M NaOH, and counted. Maximum binding observed in the absence of any cold competitor was 10%. An irrelevant GST-fusion protein was used as a negative control.

2.2. *Disulphide mapping of the six cysteines present in β_2I-5*

The whole β_2-I molecule (400 μg) was dissolved into 200 μl of 50 mM sodium phosphate (pH 7.8) and incubated for 16 h at 37°C with 10 μg endoproteinase Glu-C (V8 protease). The digested peptides plus a control of starting material were run on a Brownlee Aquapore RP300 column (C8) in 0.1% TFA and peptides were eluted with a 2–60% gradient of acetonitrile over 60 min, then 60–100% over the next 15 min. The N-termini of the main peptides were sequenced on an Applied Biosystems 470A protein sequencer with online PTH analysis, to determine whether or not pairs of disulphide-linked peptides derived from the fifth domain were present.

2.3. *Modelling of β_2I-5*

β_2I-5 was modelled by homology with the 16th domain of factor H (H16) [13]. The two sequences were aligned by hand (and taking into account the pattern of disulphide linkage reported below). The two CCP-modules share 11 identical residues in equivalent positions, in-

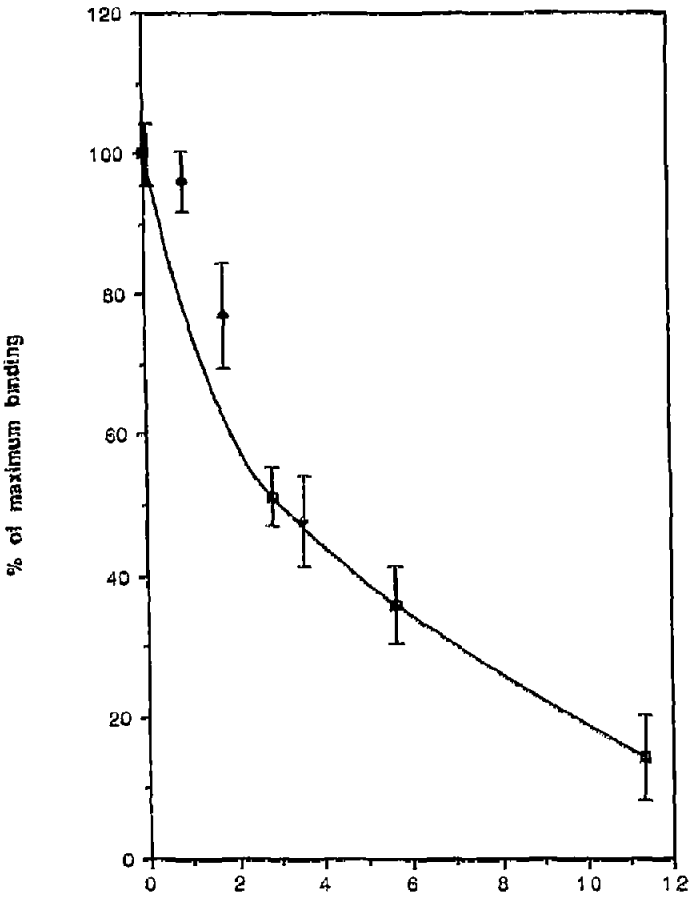

Fig. 1. Competitive inhibition of the binding of ^{125}I-labelled β_2-I to cardiolipin, coated on microtiter plates, by unlabelled β_2-I (open squares) and by the 5th domain fusion protein (β_2I-5; filled squares). The values represent the average of three measurements.

cluding the consensus cysteines, but not the highly conserved tryptophan. β_2I-5 has an insertion of 8 residues within the 'hypervariable' loop region [13], and an extension of 19 residues at the C-terminus. The modelling was done manually by amino acid replacement within the programme INSIGHT [17]. The modelled structure was energy minimised within the programme XPLOR [18]. No attempt was made to model the bulk of the C-terminal extension

3. RESULTS AND DISCUSSION

The fifth domain of human β_2-I was expressed as a GST-fusion protein in an *E. coli* expression system. After lysis and purification using glutathione agarose beads the recombinant fusion protein was used for a competition binding assay to the negatively charged phospholipid, cardiolipin. As shown in Fig. 1, the binding of radiolabelled β_2-I is inhibited by cold β_2-I and only marginally less by the GST-β_2I-5 fusion protein. An irrelevant GST-fusion protein did not have any direct binding affinity to cardiolipin (data not shown). These data indicate that the highly positively charged 5th domain, containing 15 lysines and one arginine, is involved in the binding to negatively charged phospholipids. This knowledge will be most useful in characterizing antigenic epitopes which are recognized by certain autoantibodies present in SLE patients. Efforts to understand this interaction are on-going in this laboratory.

CCP modules are normally recognised on the basis of a consensus sequence including a framework of four conserved cysteine residues disulphide linked in a 1-3 and 2-4 pattern. The first four β_2-I domains are typical examples of the CCP superfamily, whereas the fifth repeat is a very aberrant member, containing two additional cysteines and a long C-terminal tail. Lozier et al. [3] reported that the two cysteines bounding on either site of the highly positively charged region 281–288 (39–46 within the module) are disulphide linked. This information rendered alignment with other CCP module sequences, which depends strongly upon the 4 framework cysteines, very difficult and precluded any attempt at modelling the tertiary structure of β_2I-5 by homology.

In order to investigate this finding and to establish the mapping of the other four cysteines present in β_2I-5, the whole native β_2-I molecule was digested with the protease V8 and the N-termini of the major peptides were sequenced. Two peaks separated by HPLC were found to contain two peptides from β_2I-5. One contained the peptides derived from amino acids 266–285 (Cys number 2, at position 281) and 303–309 (Cys number 5, at 306). The second fraction contained the peptides 286–292 (Cys number 3, at 288) and 315–326 (Cys number 6, at 326). Since the V8-peptides were separated by reverse-phase HPLC without prior reduction, peptides eluted in the same fraction are cross-linked by disulphide bridges. Cys 2 is therefore linked to Cys 5 and Cys 3 is linked to Cys 6. Therefore, Cys 1 is linked to Cys 4 (see Fig. 2a). The two additional cysteines are thus disulphide-linked to each other and the four conserved cysteines are bridged in the same pattern as found in all other CCP modules: the disulphide bridging pattern is therefore consistent with other CCP modules.

Recently two other groups established the bovine β_2-I sequence and investigated the disulphide pattern [19,20]. They, too, concluded that the two additional cysteines of β_2I-5 are cross-linked. There is 83% amino acid identity between the human and the bovine se-

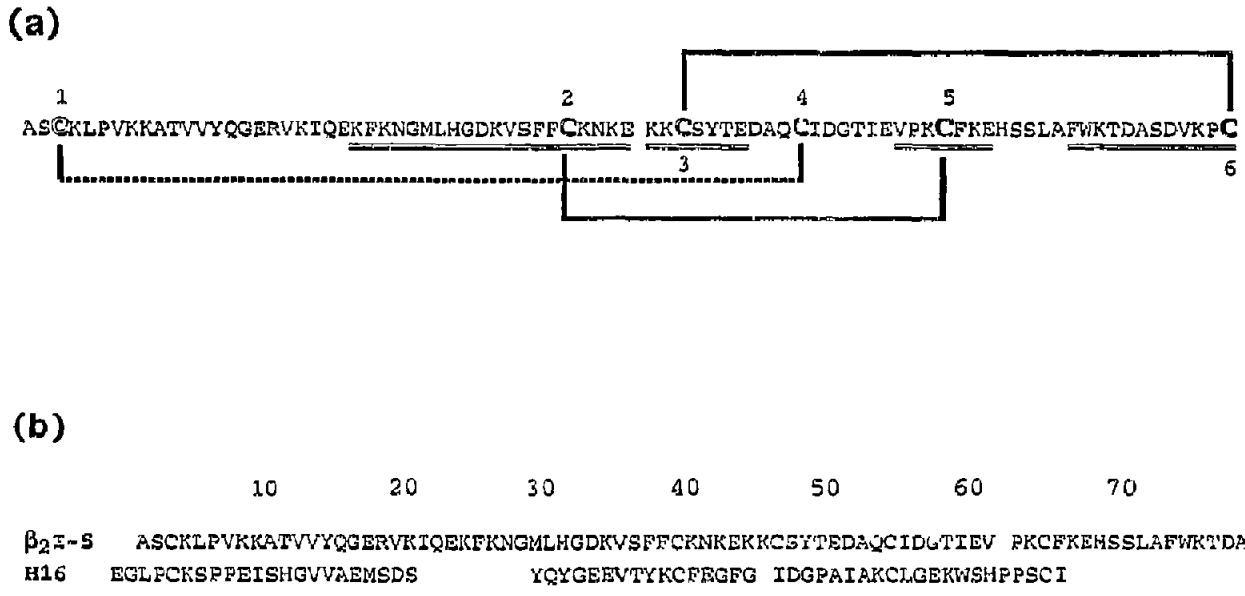

Fig. 2. (a) Disulphide mapping of the 5th β_2-I domain. V8 protease-derived peptides are double underlined. Cysteines present in peptides which were eluted in the same HPLC fraction are linked by continuous lines; Cys 2-5 and Cys 3-6. A further predicted disulphide bond between Cys 1 and Cys 4 is shown by a broken line (b) Alignment of β_2I-5 and the 16th domain of factor H (H16).

quence. Taking these reports together with the data presented here, it may be concluded that the disulphide pattern reported by Lozier et al. [3] was in error.

On the basis of sequence alignment with H16, some regions of β_2I-5 (amino acid positions 3–14, 33–43, 51–61) were expected to be in β-strands and -sheets. These were replaced without difficulty. However, the hypervariable region [13] (residues 15–28) contains 14 residues compared to 6 for H16. If it is assumed that the structurally conserved regions will stay in the same positions with regard to the rest of the module then there are only two choices of how to model this region. The 13 residues could form a large unstructured loop or they could form a helix. The choice of an α-helix seems to be more plausible (see Fig. 3). Considering the three hydrophobic residues, at positions 19 (Val), 21 (Ile) and 25 (Phe), an α-helix starting from position 17 can be arranged to orientate at least residues 21 and 25 into the core of the module. In addition, structure prediction using either Chou–Fasman (C-F) [21] or Garnier–Osguthorpe–Robson (G-O-R) [22] algorithms predicted a helix for most of that region (C-F, 17–25; G-O-R, 15–26). A critical region is that between residues 40 and 55. Here lies Cys^{46} (=Cys number 3) which is non-conserved. Residues 50–54 were assumed to be part of the conserved β-strand that has been seen in all three CCP modules for which structures are available. The position of Cys^{46} in β_2I-5 is occupied by a residue (Ile^{39}) that contributes to the hydrophobic core region in both H15 and H16. Replacement of this residue (Ile^{39}) in H16 by Cys^{46}, therefore seemed to be a reasonable option. Trp (position 52 in H16) is one of the most highly conserved residues in the CCP module sequence, apart from the cysteines. In the case of β_2I-5 it is replaced by another large hydrophobic residue, Ile^{59} (see Fig. 2b). The structural ramifications of this replacement are unknown. No attempt has been made to model the long and unconserved extension (residues 64–80) at the C-terminus of β_2I-5. There are, however, some observations that can be made regardless of the conformation of the bulk of the extension (see Fig. 3). One is that the extension leaves the molecule at residue 64, in an analogous fash-

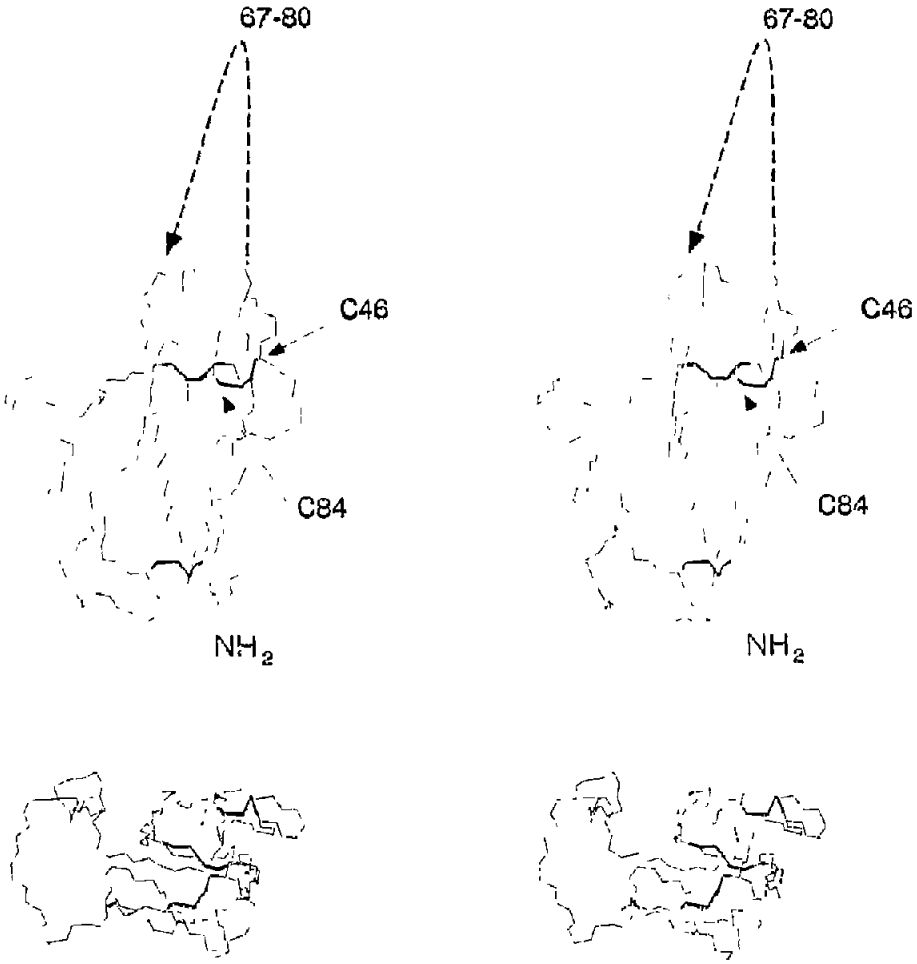

Fig. 3. (Top) Stereoview (cross-eyed) of the energy minimized model β_2I-5. The backbone is represented by a trace of the Cα and amide N positions, the disulphides by thicker lines. No attempt was made to model the large insertion containing residues 67–80, which is displayed schematically as a dotted line and arrowhead, illustrating where the loop leaves and re-enters the modelled part of the structure. The non conserved Cys^{10} and Cys^{84}, which are disulphide-linked, are indicated by arrows. (Bottom) Stereoview (cross-eyed), as above but rotated by 90° such that the N-terminus is facing away from the viewer. Residues 67–80 are not shown.

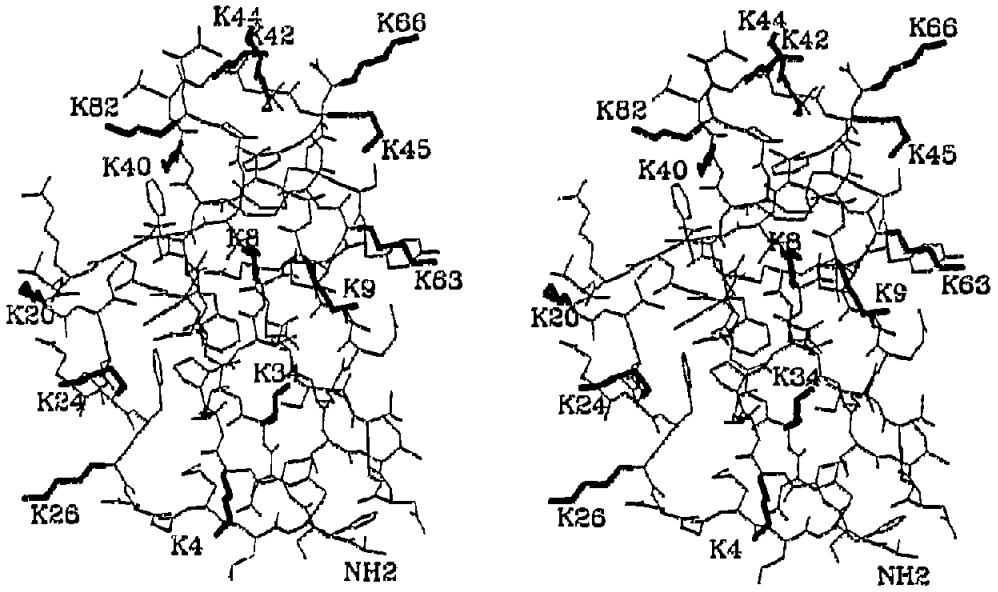

Fig. 4 A stereoview (cross-eyed) of the energy minimised model of β_2I-5, highlighting the lysine residues. All non-hydrogen backbone and side chain atoms are indicated by the thin trace, except for the side chains of the 14 lysines which are represented by thicker lines. The remaining Lys^{75} is within the insertion 67–80 which was not modelled, and is not shown here.

ion to the 'linker' region in other, non C-terminal modules. The non-conserved disulphide bridge formed between Cys^{46} and Cys^{84} (=Cys 3 and 6) will then be on the far side of the molecule with respect to residue 64 (Fig. 3). The effect is to bring a lysine residue (Lys^{82}) at the C-terminus into close proximity to the other lysine residues in the region 40–45.

Fig. 4 shows the position of 14 (out of 15) lysines in the module and indicates that a cluster of lysines, 40, 42, 44, 45, 66 and 82, lies towards the distal end of the molecule, and may form a binding region. These residues are conserved in the rat [23] and bovine sequences [19,20], except that Lys^{82} is Thr in the rat. The role of the loop 64–80 is unknown, but the presence of the hydrophobic sequence, LAFW, which is totally conserved in the rat and bovine sequences, hints at a possible interaction with the fatty acid chains of the phospholipid aggregate. The lysines mentioned above and the hydrophobic residues of the loop would obviously represent interesting targets for site-directed mutagenesis.

Acknowledgements: We would like to thank Dr. K.B.M. Reid for helpful discussions and criticism of the manuscript. This work is a contribution from the Oxford Centre for Molecular Sciences (OCMS) which is supported by SERC and the MRC. A.S was supported by an OCMS post-doctoral fellowship.

REFERENCES

[1] Polz, E. and Kostner, G.M. (1979) FEBS Lett. 102, 183–186.
[2] Schultze, H.E., Heide, K. and Haupt, H. (1961) Naturwissenschaften 48, 719.
[3] Lozier, J., Takahashi, N. and Putnam, F.W. (1984) Proc. Natl. Acad. Sci. USA 81, 3640–3644.
[4] Steinkasserer, A., Estaller, C., Weiss, E.H., Sim, R.B. and Day, A.J. (1991) Biochem. J. 277, 387–391.
[5] Steinkasserer, A., Cockburn, D.J., Black, D.M., Boyd, Y., Solomon, E. and Sim, R.B. (1992) Cytogen. Cell Genet. 60, 31–33.
[6] Kroll, J., Larsen, J.K., Loft, H., Ezban, M., Wallevik, K. and Faber, M. (1976) Biochim. Biophys. Acta 434, 490–501.
[7] Schousboe, I. (1979) Biochim. Biophys. Acta 579, 393–408.
[8] Schousboe, I. (1980) Thromb. Res. 19, 225–237.
[9] Kertesz, Z., Steinkasserer, A. and Sim, R.B. (1992) Biochim. Biophys. Acta (in press).
[10] McNeil, P.H., Simpson, R.J., Chesterman, C.N. and Krilis, S.A. (1990) Proc. Natl. Acad. Sci. USA 87, 4120–4124.
[11] Galli, M., Comfurius, P., Maassen, C., Hemker, H.C., DeBaets, M.H., van Breda-Vriesman, P.J.C., Barbui, T., Zwaal, R.F.A. and Brevers, E.M. (1990) Lancet (i) 335, 1544–1547.
[12] Reid, K.B.M. and Day, A.J. (1989) Immunol. Today 10, 177–180.
[13] Norman, D.G., Barlow, P.N., Baron, M., Day, A.J., Sim, R.B. and Campbell, I.D. (1991) J. Mol. Biol. 219, 717–725.
[14] Barlow, P.N., Norman, D.G., Steinkasserer, A., Horne, T.J., Pearce, J., Driscoll, P.C., Sim, R.B. and Campbell, I.D. (1992) Biochemistry 31, 3626–3634.
[15] Smith, D.B. and Johnson, K.S. (1988) Gene 67, 31–40.
[16] Steinkasserer, A., Solari, R., Mott, H.R., Aplin, T., Robinson, C.C., Willis, A.C. and Sim, R.B. (1992) FEBS Lett. 310, 63–65.
[17] Insight, Biosyn Inc., San Diego
[18] Brunger, A.T. (1990) XPLOR V2.2, Yale University, New Haven, CT.
[19] Kato, H. and Enjyoji, K. (1991) Biochemistry 30, 11687–11694.
[20] Bendixen, E., Halkier, T., Magnusson, S., Sottrup-Jensen, L. and Kristensen, T. (1992) Biochemistry 31, 3611–3617.
[21] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 211–222.
[22] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[23] Aoyama, Y., Chan, Y.L. and Wool, I.G. (1989) Nucleic Acids Res. 17, 6401.