Volume 250, number 1, 78–84 FEB 07262 June 1989

Review Letter

Structures and functions associated with the group of mammalian lectins containing collagen-like sequences

Steffen Thiel and Kenneth B.M. Reid[+]

Institut for Medicinsk Microbiologi, J.B. Winsløwsvej 19, 5000 Odense C, Denmark and *[+]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England*

Received 19 April 1989

The number of proteins found in body fluids and at cell surfaces, which are known to display carbohydrate-binding properties, continues to increase rapidly. In these proteins, in addition to a domain associated with lectin properties, one or more, non-lectin domains are present. It is possible that binding of sugar residues by the lectin domain may be important in triggering a variety of recognition and clearance mechanisms via the non-lectin domains. The group of lectins containing collagen-like sequences may provide some insight into structure/function relationships of the different domains in view of the well defined structures already available for several of these molecules.

Complement; Collagen-like structure; C1q; Structural comparison; Lectin; Complement lectin interaction

1. INTRODUCTION

Lectins, defined as carbohydrate-binding proteins of non-immune origin that agglutinate cells and/or precipitate polysaccharides and glycoproteins [1], were first described in 1888 when a plant lectin (a phytohaemagglutinin) isolated from castor oil seeds, was shown to agglutinate erythrocytes. At the turn of the century the first vertebrate lectin, conglutinin, was described and characterized as the component of ox serum which promoted agglutination of erythrocytes coated with activated complement components [2]. However, the difficulty of demonstrating the presence of conglutinin activity in other species resulted in this mammalian lectin attracting only limited attention. The discovery of a membrane bound lectin, the galactosyl receptor, in 1974 [3] initiated intensive research for other vertebrate lectins, which has resulted in the discovery of quite a wide variety of both membrane bound and soluble lectins (see table 1). The current interest in lectins has been stimulated by the belief that they may be involved in recognition in a variety of biological systems (reflecting the numerous oligomeric carbohydrates present on soluble as well as membrane associated proteins) and that the recognition event may in certain cases trigger effector functions.

Correspondence address: K.B.M. Reid, MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England

Abbreviations: ELAM-1, endothelial leukocyte adhesion molecule 1; GMP-140, granule membrane protein of 140 kDa; gp120, gp120 envelope protein of human immunodeficiency virus; MBP, mannan binding protein; MEL-14 antigen, mouse lymph node homing receptor; HIV-1, human immunodeficiency virus type 1; IGF-II receptor, insulin-like growth factor II receptor; SP-A, pulmonary surfactant apoprotein

2. MAMMALIAN LECTINS

Examples of the mammalian lectins which have been well characterized, are given in table 1. On

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

Table 1

Carbohydrate binding proteins with and without collagen-like domains

	Soluble			Membrane bound		
	Found in plasma or other body fluids	Sugar specificity	Lectin-type[a] 'C', 'S' or neither ('N')	Membrane bound	Sugar specificity	Lectin-type 'C', 'S' or neither ('N')
Collagen-like domain present	mannan binding protein	mannose/NADG	C	mannan binding protein (hepatic membrane associated form)	mannose/NADG	C
	lung surfactant protein	mannose/fucose	C			
	conglutinin	NADG/mannose	C			
	[C1q][b]	galactans	N			
Do not show the presence of collagen-like structure	serum amyloid P component (SAP)	MOβDG[c] and mannose terminated structures	N	galactosyl receptor	galactose/NAD gal	C
	cartilage proteoglycan core protein		C	mannose-6-P-receptors (found in a large Ca^{2+}dependent, and a small Ca^{2+} independent form)	mannose-6-phosphate	N
	fibroblast proteoglycan core protein	fucose/galactose	C			
	β-galactosidase specific lectin (Ca^{2+} independent)	β-D-galactosyl β-thiogalactopyranoside/lactose (in presence of reducing agents)	S	MEL-14 antigen	mannose-6-phosphate	C
				mannose specific lectin[d] (macrophage derived)		
				fucose binding lectin	mannose	not known
				ligatin[d]	fucose	C
					α-glucose-l-phosphate	
				complement receptor type 3	β-glucan	not known N

[a] Lectin type of 'C' or 'S' is based on the presence of certain conserved residues in the lectin domain as defined by Drickamer [4]

[b] Although there are reports that C1q has lectin activity this point is not absolutely clear especially since recent work [34] indicates that a peptide sequence in the Fc region of IgG, which is free of carbohydrate serves as an inhibitor of C1q binding

[c] MOβDG: methyl 4,6-*O*-(1-carboxyethylidene)-β-D-galactopyranoside

[d] Full structures have not yet been established for these molecules, therefore the possibility of the presence of collagen-like structure cannot be completely excluded

the basis of structural and functional studies they can be considered to fall into three general categories as defined by Drickamer [4]: the 'C-type', which show Ca^{2+} dependence and contain a 130 amino-acid-long domain which has a framework sequence of 18 conserved amino acid residues (fig.1), as seen in proteins such as the galactosyl receptor; the 'S-type', which do not show Ca^{2+} dependence but are thiol dependent and contain an approx. 100 amino-acid-long domain which has a framework sequence of 39 conserved amino acid residues (distinct from the 'C-type'), as seen in proteins such as the different β-galactoside specific lectins; a third group which does not show the presence of either C-type or S-type conserved residues and includes proteins such as serum amyloid protein and the mannose-6-phosphate receptor (which is identical to the IGF-II receptor).

The precise functions of most of these lectins after binding the appropriate carbohydrate ligands are not known, but the following roles have been proposed for several of the membrane bound lectins: elimination of aged desialylated erythrocytes and glycoconjugates showing exposed galactose groups, by the galactosyl receptor (also named asialoglycoprotein receptor) [5]; endocytosis of mannose/*N*-acetylglucosamine terminated glycoconjugates by the mannose specific lectin found on macrophages [6]; phagocytosis of yeast particles by complement receptor type 3 [7]; intracellular localization of lysosomal enzymes containing mannose-6-phosphate groups by mannose-6-phos-

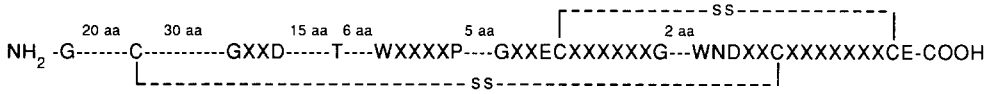

Fig.1. The conserved residues in COOH-terminal carbohydrate-recognizing domains (C-type). Positions denoted as X are variable residues within sections of constant length containing highly conserved residues. The number of amino acids found between these sections containing conserved residues is variable and in each case is indicated by --- plus the average number of amino acids found as calculated from the sequences of the proteins detailed below. The conserved residues are found in the galactosyl receptor, chicken hepatic lectin, membrane associated and serum MBP, fucose specific lectin (M_r 88-form), SP-A, flesh fly lectin, sea urchin lectin, cartilage proteoglycan core protein, fibroblast proteoglycan core protein, lymphocyte IgE FcR receptor, pancreatic stone protein and thrombomodulin, ELAM-1, GMP-140 and MEL-14 antigen [4,11–13,37,38].

phate receptors [8]. Some of the soluble lectins are also probably involved in clearance mechanisms as described below for those lectins containing collagen-like sequences and as seen with serum amyloid P component and C reactive protein which bind to a variety of polysaccharides represented on diverse bacteria, fungi and parasites [9]. β-Galactoside specific lectins are of interest in that, besides other properties, they have been shown to be mitogenic to plasma cells [10]. A new family of adhesion proteins has recently been described (ELAM-1, GMP-140 and the MEL-14 antigen) which are each characterized by containing a single C-type lectin domain, near the N-terminal end of the molecule, linked to an EGF domain and between 2 and 9 domains of the type found in the complement proteins which bind C3b and/or C4b [11–13]. These newly described adhesion proteins appear to be involved in directing leukocytes to their appropriate sites in the body.

It is possible that a clear understanding of some of the roles which lectins play after binding carbohydrate will soon be achieved from studies carried out on those lectins containing collagen-like regions since this distinctive structural feature allows the construction of accurate molecular models.

3. LECTINS CONTAINING COLLAGEN-LIKE SEQUENCES

Both the membrane associated, hepatocyte form, of mannan binding protein and the serum form have been found to contain collagen-like sequences and a C-type lectin domain, within the same polypeptide chain (table 1, fig.2). These calcium-dependent, mannose and *N*-acetyl-D-glucosamine specific lectins are designated mannan-binding proteins (MBPs) because of their specificity for yeast mannan. They exist as polymeric molecules having polypeptide chains of approx. 30 kDa. In these MBPs each chain is composed of a NH_2-terminal end followed by a collagen-like domain and finally a COOH-terminal globular domain [14] (cf. the serum form of MBP in fig.2).

It has been shown that the mannose-binding portion of the membrane associated MBP is located in the COOH-terminal non-collagen-like domain which contains the lectin C-type conserved residues [14].

The serum form of MBP has been characterised in man, cows and rabbits and is similar to the hepatic form with respect to binding specificity, pH and calcium ion dependence. They show immunochemical cross-reactivity and they are comparable in the size of their subunits, but differ from each other in the number of subunits per protein (the serum MBP being the larger of the two molecules). Human and rat soluble MBPs show approx. 60% identity over the first 20 residues of the NH_2 terminal sequences made by direct analysis at the protein level and by cDNA cloning studies [15] they show approx. 50% overall identity. The serum MBP is unusual for a plasma protein in that it has regions of collagen-like amino acid sequence (i.e. regions of repetitive -Gly-Xaa-Yaa- sequences). Two other plasma proteins, conglutinin [16] and complement component C1q [17], and also pulmonary surfactant apoprotein SP-A [18], have been shown to contain collagen-like features (fig.2). As indicated in table 1 these proteins can be classified as lectins, with the possible exception of C1q (see later) which is included primarily as a well-defined example of the structural organisation of a serum protein containing both collagen-like and globular sequences. All these proteins contain a short N-terminal, non-collagen-like sequence

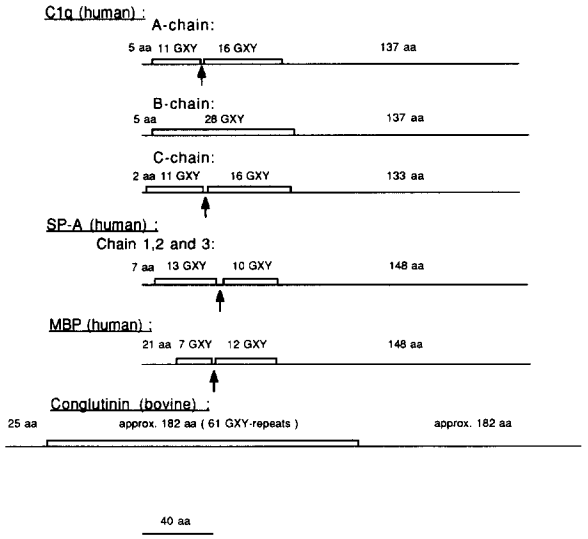

Fig.2. The polypeptide chains of C1q and of soluble lectins which contain collagen-like sequences: G-X-Y and the open box indicate the presence of Gly-Xaa-Yaa triplet repeats. (↑) Points of disruption seen in the Gly-Xaa-Yaa repeating structure.

followed by a collagenous region and a C-terminal end containing a recognition domain.

Conglutinin, found in cows [2] and humans [19], as outlined above, is the first described vertebrate lectin. It shows a calcium-dependent binding to *N*-acetyl-D-glucosamine and mannose-terminated carbohydrate structures [20]. All available evidence indicates that conglutinin binds via polymeric mannose structures to a degradation product of the major complement component C3 (this degradation product is an inactivated form of C3b which is designated iC3b) [21]. Whether this binding is involved in a physiological function of the protein is not confirmed, but it has been shown that conglutinin exhibits a leucocyte and complement-dependent bactericidal activity [22]. In this context, it is interesting that consumption of conglutinin during complement activation (bacterial infections) in cattle has been demonstrated [23]. Partial sequence studies on the bovine molecule indicates that the conglutinin polypeptide chains are composed of a 25 residue non-collagen-like N-terminal sequence followed by a 20 kDa collagen-like region and a 20 kDa globular C-terminal sequence [16] (fig.2).

SP-A, present at the surface of the lung alveoli as a component of pulmonary surfactant (which is involved in the reduction of surface tension), has been identified in several species displaying chains of 38–40 kDa and containing a lectin-domain similar to that found in other soluble lectins. A calcium-dependent binding to phospholipids [24] and carbohydrates [25] has been demonstrated for one SP-A which is composed of disulfide linked polypeptides, with the disulfide bonds being positioned in the NH_2-terminal region preceding a collagenous domain followed by a COOH-terminal non-collagen domain where ligand binding probably occurs.

C1q could tentatively be grouped with the lectins based on the reports of C1q precipitation with polysaccharides [26] and based on the ability certain lipopolysaccharides have to bind to C1q hereby mediating classical pathway activation [27]. It has been suggested that the binding of C1q to IgG_2-molecules may also involve a lectin binding, since it has been shown (using a monoclonal antibody) that classical pathway activation mediated by an antibody-antigen complex can be abolished if the antibody is non-glycosylated (achieved via treatment of the hybridoma cells with tunicamycin) [28]. However, the reports indicating that C1q may contain lectin-like binding ability have to be interpreted with caution in view of the recent publication (referred to in table 1) which shows that a short peptide sequence derived from the Fc region of IgG, and which is free of carbohydrate, serves as an inhibitor of C1q binding.

C1q has a molecular mass of 460 kDa and is composed of 18 polypeptide chains (6A, 6B and 6C) each containing a short non-collagen-like N-terminal region (2–5 residues) followed by a collagen-like region (of approx. 80 residues) and a C-terminal globular region (approx. 136 residues). The collagen-like regions form 6 triple-helices while the globular regions form 6 globular 'heads' (each 'head' being composed of approx. 136 residues derived from an A, a B and a C chain) [17]. The collagen-like regions are involved in the binding or the proenzymes C1r and C1s (the other subcomponents of the C1 complex which is the first component of the classical pathway of the serum complement system) while the C-terminal globular heads are involved in ligand binding i.e. to the Fc regions of IgG and IgM present in immune complexes. After binding of immune com-

plexes and activation of the proenzymes C1r and C1s, within the C1 complex, the activated C1r and C1s are rapidly removed by C1-inhibitor leaving the collagenous regions of C1q free to bind to cell surface C1q receptors [29].

4. STRUCTURAL SIMILARITIES BETWEEN THE COLLAGEN-LIKE SOLUBLE LECTINS AND C1q

Chemical and electron microscopy studies of SP-A [30] mannan binding protein [15] (Thiel, S. and Timpl, R., unpublished electron microscopy data) and conglutinin [16,31] have shown that they have the same distinct divisions of collagen-like and globular amino acid sequences in peptide chains between 24 and 65 kDa (fig.2) and similar overall macromolecular organisation. SP-A is difficult to distinguish from C1q when viewed in the electron microscope, the MBP also shows marked similarity to C1q but appears to be mainly found as extended tetramers and trimers rather than as a close-packed hexamer, while conglutinin is organised in a more flexible 'spider-like' structure (fig.3). All three of these proteins bind carbohydrate, the primary ligand for SP-A probably being glycolipids, while mannose binding protein and conglutinin bind mannose and *N*-acetyl-D-glucose-amine terminated structures (table 1). Protein sequence studies have shown that the non-collagen-like sequences, which form the globular 'heads' found in the SP-A, conglutinin (Thiel, S. and Willis, A.C., unpublished) and mannose binding protein, contain the 'framework' residues of C-type lectin domain [4]. Thus the carbohydrate binding is clearly mediated by the globular 'heads' in these molecules and this binding may be important in triggering a variety of recognition and clearance mechanisms via the non-lectin domain i.e. the collagen-like domain. Like C1q, the SP-A and MBPs shows a characteristic bend halfway along their collagen-like triple helical regions due to a disruption of the -Gly-Xaa-Yaa- repeating pattern of amino acids in one or more of the chains involved in the formation of the triple helices (fig.4). It is possible that this distinctive feature may be of importance in the binding of C1q to cell-surface receptors. In view of the marked similarity in shape and overall dimensions of C1q, SP-A and MBP it is not surprising that preliminary evidence has been obtained that all three can bind to the C1q-receptor (or similar receptors) found on a

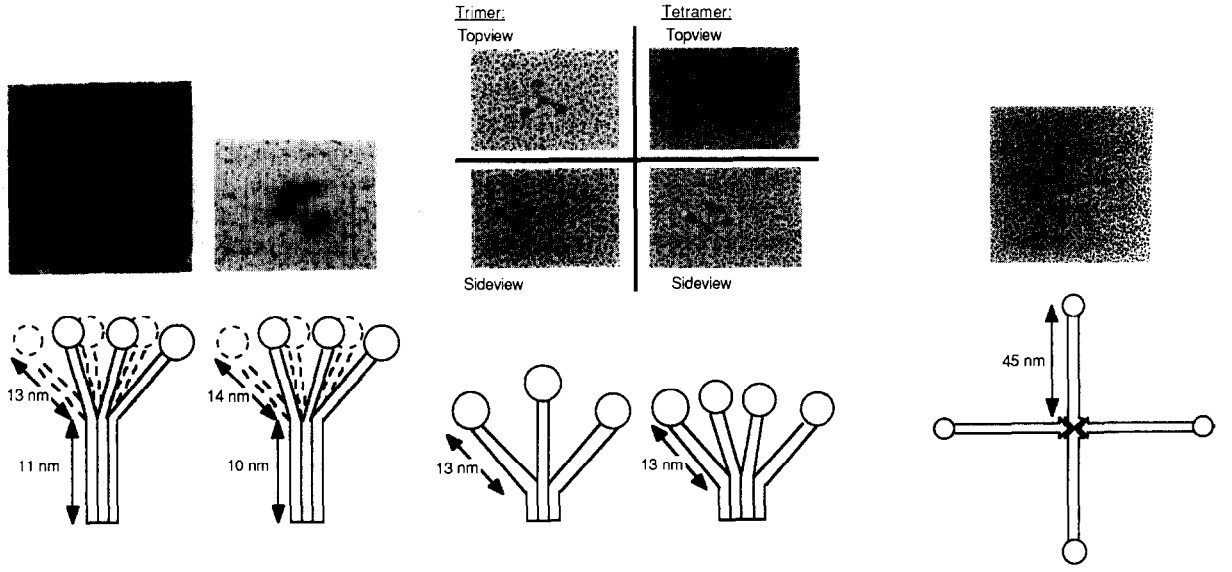

Fig.3. Electron microscopy pictures of (from left to right): C1q, SP-A, MBP (trimeric and tetrameric forms) and bovine conglutinin. It should be noted that the conglutinin molecule is considerably larger than C1q, SP-A or MBP and is shown at approx. one-third the scale used for the other molecules. Underneath the pictures the proposed structure of each protein with distances indicated (with kind permission from Dr Isliker, Dr J. Engel and Dr C.J. Strang).

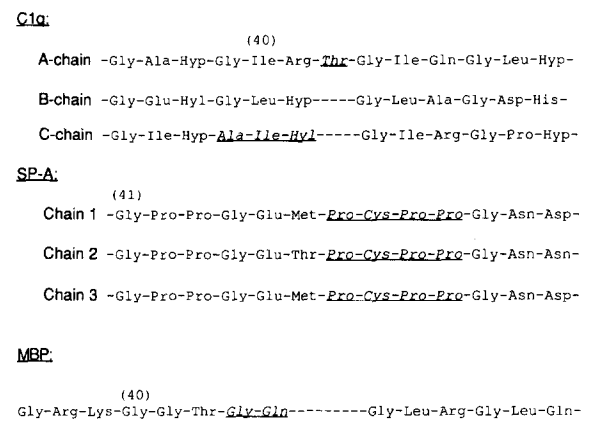

Fig.4. Amino acid sequences found around the midpoints of the collagen-like regions seen in C1q [17], MBP [15] and SP-A [18,39]. The numbers in brackets indicate the numbering of residues in each chain, with the N-terminal residue of the mature polypeptide chain in each case being designated as 1. SP-A chains 1, 2 and 3 represent the different amino acid sequences derived from cDNA [18] and genomic clones [39]. Interruptions in the repeating Gly-Xaa-Yaa sequence are underlined. (---) Space introduced in C1q B and C chain sequences and in the MBP sequence to allow alignment with the C1q A chain sequence. It is therefore predicted that any triple helix formed from these collagen-like sequences would show a disruption or 'bend' in the region of residues 40–50.

wide variety of lymphoid and other cell types [29,32]. The finding that conglutinin is involved in leucocyte mediated bactericidal effects after binding, presumably via its globular lectin domains [22], to the carbohydrate on iC3b attached to bacterial cell surfaces, is suggestive that the collagen-like regions of this molecule may also interact with leucocyte cell surface proteins of a similar nature to the C1q-receptor. Thus the binding of the collagen-like regions of these molecules to a common receptor, or to a family of receptors, could be of considerable physiological significance since this could be a general recognition and clearance mechanism for the ligands of C1q, SP-A, MBP and conglutinin, i.e. for clearance of immune complexes, glycolipids and mannose coated proteins/cells, respectively. There appears to be a further, functional, similarity between C1q and MBP since it has been reported [33] that, after binding to its ligand, the MBP can activate the classical pathway of complement thus suggesting that it may be able to mimic the role that the collagen-like region of C1q is thought to play in the activation of the proenzyme $C1r_2C1s_2$ complex.

Although there is a strong overall similarity in structure, especially over the collagen-like regions, between C1q and these lectin molecules there is no clear amino acid sequence homology between the C-terminal 'globular' domains of the chains of C1q with the C-terminal lectin domains in MBP and conglutinin, and indeed the most recent data concerned with C1q binding to its ligands suggest that this binding takes place to peptide rather than carbohydrate structures [34].

5. CONCLUSION

The lectins containing collagen-like sequences may play important recognition and protective roles at early stages of infection prior to the development of efficient humoral or cellular immunity. Recent publications [35,36] on MBP appear to support this view e.g. since it has been shown that human MBP can bind to gp120 of HIV-1 and prevent entry of the virus into $CD4^+$ cell lines [5]. Also, it has been shown that a mouse serum factor (designated Ra-reactive factor due to its specific reaction with the carbohydrate on the Ra chemotype strains of *Salmonella* and to R2 strains of *E. coli*), which appears likely to contain mouse MBP, is involved in the activation of the classical pathway of complement and resulting bactericidal effects [36]. The finding that conglutinin exhibits a leucocyte and complement-dependent bactericidal activity [22] is also indicative of the possible importance of the protective roles these lectin molecules may play during infection.

REFERENCES

[1] Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T. and Sharon, N. (1980) Nature 285, 66.
[2] Coombs, R.R.A., Coombs, A.M. and Ingram, D.G. (1961) The Serology of Conglutination and its Relation to Disease, Blackwell Sci. Publ. Oxford, Bartholomew Press, Dorking (in which primary references may be found).
[3] Ashwell, G. and Morell, A.G. (1974) Adv. Enzymol. 41, 99–128.
[4] Drickamer, K. (1988) J. Biol. Chem. 263, 9557.
[5] Weigel, P.H. (1987) in: Vertebrate Lectins (Olden, K. and Parent, J.B. eds) pp.65–91, Van Nostrand Reinhold Company Inc.

[6] Lennartz, M.R., Wileman, T.E. and Stahl, P.D. (1987) Biochem. J. 245, 705–711.
[7] Ross, G.D., Cain, J.A. and Lachmann, P.J. (1985) J. Immunol. 134, 3307–3315.
[8] Sahagian, G.G. (1987) in: Vertebrate Lectins (Olden, K. and Parent, J.B. eds) pp.46–64, Van Nostrand Reinhold Company Inc.
[9] Pepys, M.B. (1981) Lancet, 653–657.
[10] Pitts, M.J. and Yang, D.C.H. (1981) Biochem. J. 195, 435–439.
[11] Bevilacque, M.P., Stengelin, S., Gimbrone, M.R., jr and Seed, B. (1989) Science 243, 1160–1165.
[12] Johnston, G.I., Cook, R.G. and McEver, R.P. (1989) Cell 56, 1033–1044.
[13] Siegelman, M.H., Van de Rijn, M. and Weissman, I.L. (1989) Science 243, 1165–1172.
[14] Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) J. Biol. Chem. 261, 6878–6887.
[15] Ezekowitz, R.A.B., Day, L.E. and Herman, G.A. (1988) J. Exp. Med. 167, 1034–1046.
[16] Davies, A.E., ii and Lachman, P.J. (1984) Biochemistry 23, 2139–2144.
[17] Reid, K.B.M. (1983) Biochem. Soc. Trans. 11, 1–12.
[18] Floros, J., Steinbrink, R., Jakobs, K., Phelps, D., Kriz, R., Recny, M., Sulzman, L., Jones, S., Taeusch, H.W., Frank, H.A. and Fritsch, E.F. (1986) J. Biol. Chem. 261, 9029–9033.
[19] Thiel, S., Baatrup, G., Fries-Christiansen, P., Svehag, S.-E. and Jensenius, J.C. (1987) Scand. J. Immunol. 26, 461–468.
[20] Young, M.N. and Leon, M.A. (1987) Biochem. Biophys. Res. Commun. 143, 645–651.
[21] Hirani, S., Lambris, J.D. and Müller-Eberhard (1986) Biochem. J. 233, 613–616.
[22] Christiansen, P.F., Thiel, S., Svehag, S.-E., Dessan, R., Andersen, O. and Jensenius, J.C. (1988) Complement 5, 194 (Abstr.).
[23] Ingram, D.G. (1972) in: Biological Activities of Complement, pp.215–228, Karger, Basel.
[24] Hawgood, S., Benson, B.J. and Hamilton, R.L., jr (1985) Biochemistry 24, 184–190.
[25] Haagsman, H.P., Hawgood, S., Sargeant, T., Buckley, D., White, R.T., Drickamer, K. and Benson, B.J. (1987) J. Biol. Chem. 262, 13877–13880.
[26] Uhlenbruck, G., Karduck, D., Haupt, H. and Schwick, H.G. (1979) Z. Immun. Forsch. 155, 262–266.
[27] Cooper, N.R. and Morrison, E.C. (1978) J. Immunol. 120, 1862–1868.
[28] Nose, M. and Wigzell, H. (1983) Proc. Natl. Acad. Sci. USA 80, 6632–6636.
[29] Erdei, A. and Reid, K.B.M. (1988) Mol. Immunol. 25, 1067–1073.
[30] Voss, T., Eistetter, H., Schäfer, K.P. and Engel, J. (1988) J. Mol. Biol. 201, 219–227.
[31] Strang, C.J., Slayter, H.S., Lachmann, P.J. and Davis, A.E., ii (1986) Biochem. J. 234, 381–389.
[32] Malhotra, R. and Thiel, S. (1989) Biochem. J., submitted.
[33] Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. and Yamashina, I. (1987) J. Biol. Chem. 262, 7451–7454.
[34] Duncan, A.R. and Winter, G. (1988) Nature 332, 738–740.
[35] Ezekowitz, R.A.B., Kuhlman, M., Groopman, J.E. and Byrn, R.A. (1989) J. Exp. Med. 169, 185–196.
[36] Ji, Y.-H., Matsushita, M., Okada, H., Fujita, T. and Kawakami, M. (1988) J. Immunol. 141, 4271–4275.
[37] Patthy, L. (1988) Biochem. J. Lett. 202, 689–697.
[38] Patthy, L. (1988) J. Mol. Biol. 253, 309–311.
[39] White, R.T., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B. and Cordell, B. (1985) Nature 317, 361–363.