Morishita, R., Kato, K., & Asano, T. (1988) *Eur. J. Biochem. 174*, 87-94.
Nukada, T., Tanabe, T., Takahashi, H., Noda, M., Haga, K., Haga, T., Ichiyama, A., Kangawa, K., Hiranaga, M., Matsuo, H., & Numa, S. (1986) *FEBS Lett. 197*, 305-310.
Oinuma, M., Katada, T., & Ui, M. (1987) *J. Biol. Chem. 262*, 8347-8353.
Schaffner, W., & Weissmann, C. (1973) *Anal. Biochem. 56*, 502-514.
Skoog, W. A., & Beck, W. S. (1956) *Blood 11*, 436-454.
Sternweis, P. C., & Robishaw, J. D. (1984) *J. Biol. Chem. 259*, 13806-13813.
Suki, W. N., Abramowitz, J., Mattera, R., Codina, J., & Birnbaumer, L. (1987) *FEBS Lett. 220*, 187-192.
Sullivan, K. A., Liao, Y.-C., Alborzi, A., Beiderman, B., Chang, F.-H., Masters, S. B., Levinson, A. D., & Bourne, H. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6687-6691.
Van Meurs, K. P., Angus, C. W., Lavu, S., Kung, H.-F., Czarnecki, S. K., Moss, J., & Vaughan, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3107-3111.

Disulfide Bonds Are Localized within the Short Consensus Repeat Units of Complement Regulatory Proteins: C4b-Binding Protein†

Jarmila Janatova,*,‡ Kenneth B. M. Reid,§ and Antony C. Willis§

Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah 84132, and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.

Received April 14, 1988; Revised Manuscript Received February 7, 1989

ABSTRACT: Several plasma and membrane proteins belong to a superfamily of structurally related proteins that contain internal homology of a variable number (2–30) of repeating units. Each SCR (short consensus repeat) unit is approximately 60 amino acid residues in length, with the positions of 1 Trp, 2 Pro, and 4 Cys residues being conserved. The aim of this study was to provide experimental evidence that each SCR may exist as an independent structural domain maintained by disulfide bonds. The well-characterized C4b-binding protein (C4BP) with eight SCR units in each of its seven identical chains was chosen for this study. Analysis of the disulfide-bonding pattern indicated that intrachain disulfide bonds may be localized within each SCR unit, with the first and third and the second and fourth half-cystines in each unit being linked. This pattern of disulfides may confer to C4BP (and to other structurally related proteins) a conformation which apparently allows the assembly of the SCR units (4–30) in a tandem fashion. Such an arrangement of the polypeptide chain(s) may explain, in part, the elongated shape of these protein molecules. The structural motif of the SCR units of C4BP is discussed in relation to those previously described for the type II domain of fibronectin and the kringle structure present in various proteins of the coagulation system.

Activation of C3, the central component of the complement system, is under the regulation of at least six proteins. These include plasma proteins factor H (Whaley & Ruddy, 1976) and C4b-binding protein (Fujita et al., 1978), and four membrane-bound proteins: receptor CR1[1] (Fearon, 1980; Medof et al., 1982), receptor CR2 (Moore et al., 1987; Weis et al., 1988), decay accelerating factor (Nicholson-Weller et al., 1982; Kinoshita et al., 1986), and membrane cofactor protein (Lublin et al., 1988). All six proteins are genetically linked in man, being under control of the RCA (regulator of complement activation) locus on human chromosome 1 (Rodriguez de Cordoba et al., 1985, 1988; Rey-Campos et al., 1988; Carroll et al., 1988; Lublin et al., 1988). They also share many features with respect to their biochemical modes of action (Holers et al., 1985).

Each of these proteins is largely composed of a varying number of short consensus repeat (SCR) units, approximately 60 amino acid residues long, in which the positions of 1 tryptophan, 2 prolines, 4 half-cystines, and some other residues are conserved (Reid et al., 1986). The SCR is found also in several other proteins (Table I), although it does not always constitute a large fraction of the total protein sequence. Data on factor B (Campbell et al., 1984), IL-2 receptor (Leonard et al., 1985), C4b-binding protein (Lintin et al., 1987), and factor H (Vik et al., 1988) suggest that each SCR unit is

† This work was supported by the U.S.–U.K. Cooperative Science Program from the National Science Foundation (Project INT-8611747), by the U.K. Medical Research Council, by two Travel Research grants from the Burroughs Wellcome Research Fund, and by funds from the Dean's Office, University of Utah School of Medicine.

* Correspondence should be addressed to this author.
‡ University of Utah School of Medicine.
§ University of Oxford.

[1] Abbreviations: SDS, sodium dodecyl sulfate; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; NEM, *N*-ethylmaleimide; DTT, dithiothreitol; MSH, β-mercaptoethanol; TFA, trifluoroacetic acid; HPLC, high-pressure liquid chromatography; FPLC, fast protein liquid chromatography; Gdn·HCl, guanidine hydrochloride; C4BP, C4b-binding protein, a cofactor for cleavage of the C4bα-chain; factor H and/or CR1, cofactors for factor I mediated cleavages of the C3bα-chain; CR1, a C3b receptor on erythrocytes and other blood cells; CR2, a C3d receptor on B-lymphocytes; DAF, decay accelerating factor present on membranes of blood and tissue cells; MCP, membrane cofactor protein; B and C2, proenzymes in the complement system; C1r and C1s, subcomponents of the first component of complement; β_2I, β_2-glycoprotein I; IL-2(R), receptor for interleukin 2; factor XIII, plasma transglutaminase; SCR, short consensus repeats in C4BP and other related proteins.

0006-2960/89/0428-4754$01.50/0

Table I: Superfamily of Structurally Related Proteins Containing Repeats of 60 Amino Acid Residues

4	7	30	32	35	46	52	57	59
---Cys-	-Pro------	-Tyr-	-Cys-	-Gly------	-Cys-----	-Trp----	-Pro-	-Cys-
		Phe					Ala	

protein	M_r ($\times 10^{-3}$)	no. of amino acids	no. of repeats	ref
Complement Regulatory Proteins				
C4BP	70 × 7	549 × 7	8	Chung et al. (1985)
H	155	1213	20	Kristensen & Tack (1986) Ripoche et al. (1988)
DAF	70	347	4	Medof et al. (1987)
MCP	45–70	350	4	Lublin et al. (1988)
CR1	190	1998	30	Klickstein et al. (1988)
CR2	145	1067	16	Moore et al. (1987)
		1012	15	Weis et al. (1988)
Complement Enzymes				
B	92	739	3	Morley & Campbell (1984)
C2	108	732	3	Bentley (1986)
C1r	83	705	2	Leytus et al. (1986) Journet & Tosi (1986)
C1s	83	720	2	Mackinnon et al. (1987) Tosi et al. (1987)
Noncomplement Proteins				
factor XIII	80	641	10	Ichinose et al. (1986)
β_2I	50	326	5	Lozier et al. (1984)
IL-2 (R)	55	247	2	Leonard et al. (1985)

usually encoded by a separate exon.

Disulfide bonds can contribute significantly to the specificity and stability of functional forms of proteins, and, in general, positions of half-cystine residues participating in inter- and intrachain bonds remain invariant in homologous proteins (Thornton, 1981; Richardson, 1981). It has been presumed that the cysteinyl residues in the complement regulatory protein factor H (DiScipio & Hugli, 1982), and other structurally related proteins (Reid et al., 1986), may be involved in a regular pattern of intrachain disulfide bonds. The aim of the present study was to test the hypothesis that in the complement regulatory proteins each SCR unit may exist as a distinct structural domain maintained by disulfide bonds.

The C4b-binding protein (C4BP), that was chosen for the study, consists of seven identical chains (M_r ~70K) that are held together by interchain disulfide bonds at their C-termini (Dählback et al., 1983; Dählback & Müller-Eberhard, 1984; Chung & Reid, 1985). Each chain contains eight SCR units, which makes the assignment of disulfides reasonably approachable. The whole chain (8 SCR units plus C-terminal non-SCR region) contains a total of 32 conserved and 4 nonconserved cysteinyl residues (Chung et al., 1985a). Images obtained by electron microscopy suggest that the C4BP molecule has a spiderlike structure with seven flexible tentacles (approximately 330 Å long) attached to a small central body, the core (Dählback et al., 1983). Analysis of the C4BP protein in solution by X-ray scattering indicated, however, that the arms of the C4BP heptamer seem to be close together (Perkins et al., 1986).

A limited fragmentation of the C4BP molecule has been achieved with several enzymes (Chung & Reid, 1985). In particular, treatment of native C4BP with α-chymotrypsin can generate two major cleavage products, with separation of the disulfide-bonded core from the tentacle (Dählback & Müller-Eberhard, 1984; Chung et al., 1985b). The tentacle is a polypeptide of M_r 48K–50K from the N-terminal part of each of the seven identical chains and comprises SCR units I–VII. The core consists of seven disulfide-linked polypeptides of M_r ~20K, each comprising SCR unit VIII and the non-SCR carboxy-terminal part of the chain.

To determine the disulfide-bonding pattern, we generated the disulfide-bonded core and the large fragments of the tentacle with α-chymotrypsin. The purified large fragments were then digested with trypsin followed by V8 protease, and the generated peptides were resolved by reverse-phase HPLC. Isolation and characterization of several disulfide-bonded peptide pairs from the C4BP chain indicated that intrachain disulfide bonds may, indeed, be localized within each SCR unit with the first and third and the second and fourth half-cystines in each unit being linked. Preliminary reports of these studies in the form of abstracts have been presented (Janatova et al., 1987, 1988).

Experimental Procedures

Materials. All chemicals and reagents used were of the highest grade available, or as appropriate. Sources of the enzymes were as follows: β-trypsin, TPCK-treated (Worthington Biochemical Corp.); *Staphylococcus aureus* protease V8 (Miles Laboratories, Ltd.); α-chymotrypsin from bovine pancreas, TLCK-treated (Sigma Chemical Co.). Gdn·HCl, Aristar grade from BDH, was used for dissolving peptides.

Isolation of C4BP. Since C4BP coprecipitates with C1, IgM, properdin, and some other proteins during the euglobulin precipitation of serum, the procedure originally described by Gigli et al. (1976) for purification of C1 was used as the basis for the isolation of C4BP.

To prepare serum, fresh-frozen plasma (CPD-A1) was thawed and made 20 mM in $CaCl_2$, and after 4 h at room temperature, the fibrin clot was separated at 4 °C by centrifugation at 12210*g* for 30 min. The serum was made 1.6 mM in DFP, and the "euglobulin" proteins were precipitated in the cold (after a 5-fold dilution of serum with water) at pH 7.4 in the presence of 5 mM $CaCl_2$, 2.5 mM iodoacetamide, and 0.2 mM phenanthroline. After 18 h at 4 °C, the precipitate was washed at pH 5.5 with 40 mM sodium acetate/5 mM $CaCl_2$. The proteins were then solubilized in the pH 5.5 buffer (50 mM sodium acetate/5 mM $CaCl_2$/200 mM NaCl) and fractionated by gel filtration in the same pH 5.5 buffer on Sepharose 6B (5 × 81 cm; 46 mL/h; 11 mL/tube). The majority of C4BP was eluted before the C1 component and properdin. The C4BP-enriched pool was dialyzed and then fractionated at pH 7.4 (10 mM NaH_2PO_4/47 mM Na_2HPO_4/5 mM Na_2EDTA) on a DEAE-Sephacel column (2.5 × 32 cm; 29 mL/h; 5 mL/tube) using a linear gradient of NaCl from 0 to 300 mM (300 mL of each).

The C4BP-containing pools obtained by ion-exchange chromatography on DEAE-Sephacel were dialyzed against the pH 5.2 buffer (50 mM sodium acetate/1 M NaCl/5 mM Na_2EDTA) in order to promote the dissociation of C1r polymers and thus to enhance the separation of C1r from C4BP by gel filtration on Sephacryl S-300 (2.5 × 85 cm; 25 mL/h; 3.5 mL/tube). Next, immunoaffinity chromatography (IA-anti-IgM; 2.5 × 10.5 cm; 10 mL cm^{-2} h^{-1}; 2.7 mL/tube) at pH 7.4 (50 mM sodium phosphate/250 mM NaCl/5 mM Na_2EDTA) was used to separate the majority of IgM from C4BP. A small amount of IgM that remained associated with C4BP (Villiers et al., 1981) was removed after incubation of

the C4BP sample with α-chymotrypsin by FPLC on a Superose 12 column, since IgM is not cleaved by α-chymotrypsin under the given conditions (see Figure 1). Ultrafiltration over a PM10 Diaflo membrane was used for concentration of protein solutions.

SDS–Polyacrylamide Gel Electrophoresis. Protein fractions obtained at various chromatographic steps during the isolation of C4BP, and the degree of C4BP fragmentation following the incubation with α-chymotrypsin, were evaluated by SDS–PAGE in a 10% (w/v) resolving gel under both nonreducing and reducing conditions. Buffers and gels were prepared as described by Laemmli (1970). Samples were prepared, and electrophoresis was carried out much as described by Janatova (1986).

Two-dimensional SDS–polyacrylamide gel electrophoresis was used to establish the relationship between unreduced and reduced fragments of C4BP and was performed as described earlier (Janatova, 1986).

Fragmentation of C4BP with α-Chymotrypsin. To establish the optimal conditions for generation of the core and fragments of the tentacle, C4BP (0.5 mg/mL; 50 mM Tris-HCl/150 mM NaCl/1 mM Na_2EDTA) was incubated with α-chymotrypsin at two different substrate/enzyme ratios (100/1; 50/1; w/w) at pH 8.0 and 37 °C for 1, 2, 3, and 5 h. Aliquots (25 μL) were withdrawn at each time point, treated with 0.2 M DFP (1 μL), and analyzed by SDS–PAGE. The largest amount of a 50-kDa polypeptide (tentacle) was present after a 1-h incubation at a substrate/enzyme ratio = 100/1. At the higher enzyme concentration and 5 h at 37 °C, almost all of the tentacle was cleaved to smaller fragments. The latter conditions were employed for generation and isolation of the C4BP fragments.

Purification of C4BP Fragments. The C4BP fragments generated by chymotrypsin cleavage (see above) were first fractionated by FPLC on a Superose 12 column (Figure 1). They were then characterized and purified by reverse-phase HPLC (Mahoney & Hermodson, 1980). Samples (200–500 μL) from peaks 2, 5, 6, 7, 8, and 9 through 12, were applied directly to a C_{18} column (3.9 × 300 mm; Partisil 10, ODS-3 Whatman, 10 μm; end-capped) preequilibrated with 4% CH_3CN in 0.1% TFA and connected to the Waters HPLC system. Peptides were eluted at a flow rate of 1 mL/min with a linear gradient of CH_3CN from 4 to 72% in 0.1% TFA. The effluent was collected manually, and pools were made according to the optical density at 206 nm.

Generation and Isolation of Disulfide-Bonded Peptides and Their S-Carboxymethylated Constituents. A sample (2.5 mL) containing ~18 nmol of peptides from *peak 6* (Figure 1B) in 0.2 M NH_4HCO_3 was incubated with 46 μg of TPCK–trypsin in the presence of 0.25 mM iodoacetamide at 37 °C. After 4.25 h, a solution of V8 protease (55 μg in 138 μL) was added, and the second enzymatic digestion was allowed to proceed at 37 °C for 12.75 h. Generated peptides were fractionated by reverse-phase HPLC (described above and in the legend to Figure 3) in three identical runs (850 μL/run). As controls, the same amounts of enzymes were incubated under the same conditions as above: either TPCK–trypsin alone, V8 alone, or V8 that was added to the TPCK–trypsin solution after 4.5 h. V8, which does not contain any disulfides (Drapeau et al., 1972; Houmard & Drapeau, 1972), yielded only one peak upon HPLC of any of the samples. All trypsin was autolyzed after 4.25 h; the level of tryptic peptides was low and did not interfere with the analysis of C4BP peptides.

To detect the *presence of cystine-containing peptides*, aliquots from each fraction (100–150 μL) were dried in glass tubes (acid-washed, furnaced) in a Speed-Vac (Savant Instruments). To dissolve peptides, 5 μL of CH_3OH and then 25 μL of 99% formic acid were added, and samples were kept on ice. To oxidize cystines, 75 μL of performic acid (prepared by incubating 0.15 mL of 30% H_2O_2 with 2.85 mL of 99% formic acid, 2 h, 22 °C) was added to each tube, and the reaction was allowed to proceed for 2 h at −3 to −5 °C. Each sample was then diluted with 400 μL of cold water and dried down. To detect the presence of cysteic acid, samples were hydrolyzed with 200 μL of Pierce constant-boiling HCl in the presence of 2 μL of phenol, and hydrolysates were analyzed as phenylthiocarbamoyl derivatives (Heinrikson & Meredith, 1984).

The pools of *cystine-containing fractions* from three runs (~18 nmol) were dried down and dissolved in 200 μL of 6 M Gdn·HCl/0.4 M Tris-HCl/2 mM Na_2EDTA (pH 8.0). After 18 h at room temperature, each sample was reduced with 20 mM DTT (25 μL of 180 mM DTT/sample) for 1.5 h at 37 °C, and each reduced fraction was alkylated with 40 mM iodoacetic acid (25 μL of 400 mM iodoacetic acid/sample) for 1 h at room temperature in the dark. Reduced and alkylated samples were either used immediately or kept frozen at −20 °C until used.

To isolate *S-carboxymethylated constituents from disulfide-bonded pairs*, each reduced/alkylated sample (250 μL) was subjected to reverse-phase HPLC under the original conditions of separation (e.g., lower panel in Figure 3). After sample application, the column was washed with 16 mL of 4% CH_3CN/0.1% TFA to remove the reagents. Peptides were then eluted with a linear gradient of CH_3CN as described in the legend to Figure 3.

N-Terminal Sequence Analysis. S-Carboxymethylated peptides were sequenced by using automated Edman degradation in an Applied Biosystems 470A protein gas-phase sequencer (Hewick et al., 1981).

Determination of Protein Concentrations. Concentration of protein solutions was determined by measuring the absorbance at 280 nm, and using absorption coefficients of 9.3 for C4BP (Villiers et al., 1981), 14.3 for β-trypsin (Decker, 1977), and 10.0 for α-chymotrypsin. In the absence of information (Drapeau et al., 1972), a value of 10.0 was also used for V8.

RESULTS

Isolation of C4BP Fragments Generated with α-Chymotrypsin. The separation of the disulfide-bonded *core* from the *tentacle* and its fragments under nondissociating and nonreducing conditions was achieved on a Superose 12 column equilibrated with 0.2 M NH_4HCO_3 using a Pharmacia FPLC system. Moreover, the separation of the C4BP fragments from contaminating IgM was also achieved in this step. Two examples of such a fractionation are illustrated in Figure 1. Individual pools were characterized by SDS–PAGE, and pools from peak 2, 5, and 8 (Figure 1A) were subjected first to reverse-phase HPLC and then to N-terminal sequence analysis. Identification of the proteins or polypeptides isolated from the C4BP digest by FPLC (Figure 1) is described below and summarized in Figure 2.

Characterization of C4BP Fragments by SDS–PAGE, HPLC, and N-Terminal Sequence Analyses. Sequence analysis of the pool from *peak 2* (Figure 1A) containing the disulfide-bonded core yielded two sequences in equal yields. Since the analysis was performed on a nonreduced sample, these results suggested that a small peptide with the sequence *KEEIIYEC*$_{406}$DKGY is linked by a disulfide bond to the core that starts with the sequence of *SAPAPQC*$_{432}$*K*ALCRKPE.... The cysteinyl residues in these sequences represent the second

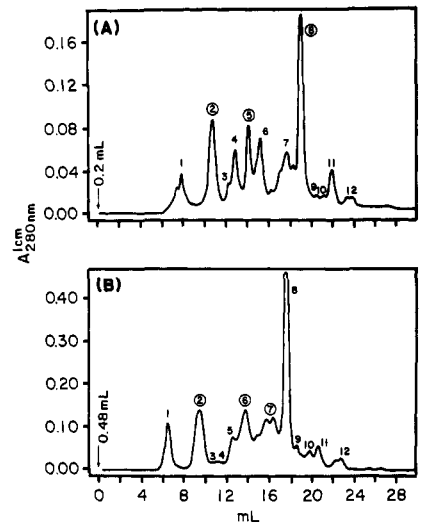

FIGURE 1: Isolation of C4BP fragments by FPLC. Human C4BP (1 mg/mL) was incubated at pH 8 (100 mM Tris-HCl/300 mM NaCl/2 mM Na_2EDTA) with α-chymotrypsin (S/E = 50/1) at 37 °C for either 5 h (panel A) or 14 h (panel B). The C4BP digest (0.5 mg in 200 μL) was spun and applied to a Superose 12 column (equilibrated with freshly made 0.2 M NH_4HCO_3) at a flow rate 0.2 mL/min (panel A). A similar separation was obtained at 0.4 mL/min and 2.4-fold larger sample volume containing 1 mg of C4BP digest (panel B). The elution profile was monitored at 280 nm, and fractions of 1 mL (0.5 mL/tube at peaks) were collected. All pools from the individual peaks were analyzed by SDS–PAGE, and pools 2, 5, and 8 (profile A) were further purified by RP-HPLC and then subjected to N-terminal sequence analyses (see text). On prolonged incubation (14 h), polypeptides from peaks 3, 4, and 5 were fragmented further, and the cleavage products were eluted within peak 6 (profile B). Results of further fragmentation of peak 6 with TPCK–trypsin, followed with V8 protease, are presented in Figure 3. Peptides from peak 7 (profile B), without any further enzyme treatment, were chromatographed on a C_{18} column.

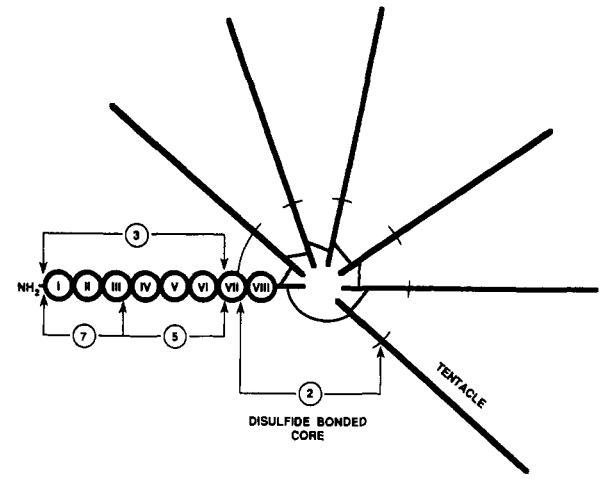

FIGURE 2: Schematic representation of C4BP and characterization of its fragments. Arrows denote sites of chymotryptic digestion. The proteins or polypeptides in the individual peaks obtained by FPLC from the C4BP digest (Figure 1) were identified as follows: peak 1, contaminating IgM; peak 2, disulfide-bonded core [M_r 140K/(7 × 20K)]; peak 3, tentacle (M_r 42K/49K); peak 4, fragments of the tentacle (M_r 32K–34K/33K and 38K); peak 5, middle of the C4BP chain (M_r 20K/24K); peak 6, several fragments from peak 5, approximately 10–11 kDa; peak 7, small peptides, approximately 2 kDa; peak 8, KQSSSY, a peptide from the SCR unit VII; peaks 9–12, very small peptides, such as SFF or SHW from the SCR unit VII. Values for molecular weight are given for the nonreduced/reduced state. While each of the seven identical chains of C4BP is 330 Å long and 30 Å in diameter, the length of the tentacle (peak 3) comprising approximately 6.5 SCR units was shown to be 290 Å (Dählback et al., 1983).

and fourth cysteines within the SCR unit VII, but it is not directly established that these two are bonded to each other.

The peptide material from *peak 5* (Figure 1A) yielded a single band of M_r 20K using SDS–PAGE under nonreducing conditions. After reduction, the apparent size was increased to 24 kDa, and the presence of small peptide(s) at the boundary was detected. One major peak was observed on purification of unreduced peptide by HPLC. Sequence analysis of this material identified two cleavage sites. The data showed that a small peptide, $SC_{155}DPRF$, with the second Cys of the SCR unit III is attached to a polypeptide of 20/24 kDa (the middle of the C4BP chain). The latter one starts with the sequence $RPSPPTC_{186}EKIT$, which contains the fourth Cys of the same SCR unit. On a prolonged incubation of C4BP with α-chymotrypsin (14 h), cleavage products of the tentacle were eluted within the *peak 6* from Superose 12 (compare Figure 1A and Figure 1B).

A relatively large number of peptides were present in *peak 7* and the adjacent earlier eluting peaks (Figure 1B). The majority of the peptides in this pool was assumed to have originated from the SCR units I and II of the C4BP chain. The peptide in *peak 8* (Figure 1A) was identified, following the rechromatography by HPLC, as having the sequence *KQSSSY* that is located within the SCR unit VII. Thus, this sequence identified the fifth and the sixth of several cleavages by α-chymotrypsin occurring within the SCR unit VII, the area between the tentacle and the core.

Isolation and Identification of Disulfide-Bonded Peptides. Further fragmentation of polypeptides from the middle section of the C4BP chain (*peak 6* in Figure 1B) was achieved by employing two additional enzymes with different specificities, TPCK–trypsin followed by protease V8 (Figure 3, top panel). The cystine-containing fractions were reduced, alkylated, and rechromatographed on the same column under the original elution conditions. The presence of S-carboxymethylated cysteinyl residues was confirmed by PICO.TAG amino acid analysis, and the *S*-carboxymethyl positive peptides were then subjected to N-terminal sequence analysis on a gas-phase sequencer. The result of one of the experiments is presented in the lower panel of Figure 3. As expected, each of the constituents from a disulfide-bonded peptide present in fraction 2 eluted at a different concentration of acetonitrile which was distinct from the concentration at which the original peptide eluted.

The results from the analyses of all the, so far, identified disulfide-bonded peptides are depicted in Figure 4 (the sequence data obtained in the present study are underlined). A peptide pair from fraction 1 was identified as $SC_{92}SE$–$SHPLPQC_{122}E$, and it contains the second and fourth Cys from the SCR unit II. Fraction 2 (lower panel in Figure 3) was shown to contain the peptide pair $C_{127}KPPPDIR$–$GHASISC_{169}TVE$, comprising the first and third Cys from the SCR unit III. The peptide VVGTVLR (SCR unit V) was isolated from fraction 3 and sequenced, while the SCMC peptides were presumably blocked. During the analyses of some other Cys-positive samples (e.g., fraction 4), no sequence data were obtained possibly due to blocked N-termini. The peptides $SAIC_{361}QGDGTW$ and $ALC_{316}C_{317}PEPK$, comprising the third and first Cys from the SCR unit VI, were separated after reduction and alkylation of peptides from fraction 5. The latter one was present in the same fraction as the peptide $SRPANHC_{339}VY$ (SCR unit VI) which contains one of four nonconserved Cys in C4BP. Either the two peptides remained linked to each other or their elution conditions were identical. The data suggest that the residue C_{339} forms an extra disulfide bridge with C_{317}. Fraction 6 yielded the

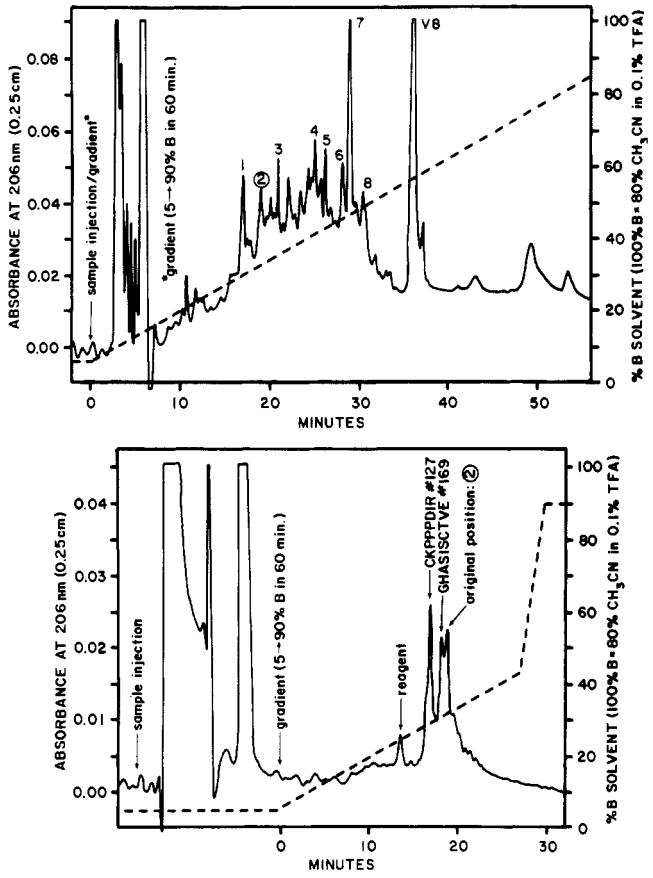

FIGURE 3: Isolation of disulfide-bonded peptides and their S-carboxymethylated constituents. A graph in the top panel represents an example of the peptide map obtained by reverse-phase HPLC of the pool from peak 6 (Figure 1B) after its incubation with TPCK–trypsin that was followed by V8. The digest was directly applied to a C_{18} column equilibrated with 4% CH_3CN in 0.1% TFA, and the chromatogram was developed with a linear gradient of CH_3CN from 4 to 72% (containing 0.1% TFA) in 60 min at a flow rate of 1 mL/min. The presence of disulfides was detected by amino acid analysis following the oxidation of cystines by performic acid. The cystine-containing fractions were dried down, dissolved in 6 M Gdn·HCl (pH 8.0), reduced with 20 mM DTT, and alkylated with a 2-fold excess of iodoacetic acid. The elution profile in the lower panel shows the results of the rechromatography of fraction 2, one of the treated fractions. As expected, the two constituents from the disulfide-bonded peptide pair, $C_{127}KPPPDIR$ and $GHASISC_{169}TVE$, eluted at different concentrations of CH_3CN, that also differed from the original eluting position. These peptides comprise the first and third Cys in SCR III. The same methods were used to identify other cystine-containing peptide pairs in fraction 1, $SC_{92}SE$–$SHPLPQC_{122}E$ (second and fourth Cys, SCR II), and fraction 5, $SAIC_{361}QGDGTW$–$ALC_{316}C_{317}PEPK$ (third and first Cys, SCR VI). The latter peptide either remained linked to or just coeluted with $SRPANHC_{339}VY$ which contains one of four nonconserved Cys in C4BP. The residue C_{339} forms apparently with C_{317} an extra disulfide bridge in the C4BP chain. Fraction 6 yielded the peptide $TPSC_{374}GDIC_{378}NFP$PK comprising the fourth Cys from the SCR VI and the first Cys from the SCR VII.

peptide $TPSC_{374}GDIC_{378}NFP$PK comprising the fourth Cys from the SCR unit VI and the first Cys from the SCR unit VII. There were technical problems with the recovery of the small peptides SC_{351}HE (second Cys from SCR unit VI) and SC_{420}SY (third Cys from SCR unit VII). Our earlier sequence data on two nonreduced samples, the disulfide-bonded core (peak 2, Figure 1A) and the 20/24-kDa polypeptide (peak 5, Figure 1A), indicated a possibility of linkages between the second and fourth Cys in the SCR units VII and III.

Further investigation of the α-chymotrypsin-generated peptides from *peak 7* (Figure 1B) by two-dimensional reverse-phase HPLC revealed polymorphism in the sequence adjacent to the third Cys from the SCR unit II. After reduction and alkylation, three "new" peptides rather than two were observed. The peptide $KRC_{65}RHPGL$ which contains the first Cys from the SCR unit II was linked to the peptide $SRC_{106}EVQDRGV$, that was predicted from the cDNA sequence (Chung et al., 1985a), as well as to another peptide with a similar composition. Its sequence, $SC_{106}DEVQDRGV$, indicated that a double mutation had taken place: deletion of Arg before, and insertion of Asp after, Cys_{106}. Other polymorphic sites in C4BP have also been identified (Chung et al., 1985b).

A total of seven disulfide-bonded pairs were localized within the sequence of C4BP (Figure 4). In addition to a complete assignment of disulfide bridges in the SCR units II and III, the identification of peptide pairs within the SCR units VI and VII further supports the original hypothesis that the first and third and the second and fourth cysteinyl residues within each SCR unit may be linked, as illustrated in Figure 5.

Discussion

Only recently has the existence of the superfamily of the functionally and/or structurally related proteins with high half-cystine and proline contents become apparent [reviewed in Holers et al. (1985) and Reid et al. (1986)]. These proteins contain internal homology of several repeating units (2–30), in which the positions of 2 Pro, 1 Trp, 4 Cys, and several other residues are conserved. It has been assumed that these repeated cysteine sequences and corresponding disulfide bond interactions may play an important role in the secondary structure that would be common to all proteins of this superfamily (Table I), including C3b/C4b-binding proteins and several noncomplement proteins.

β_2-Glycoprotein I is the first protein in which the existence of the 60 amino acid residue long repeating units has been recognized. Although it has been suggested "that certain homologous sets of cysteines are linked to other homologous sets of cysteines" in β_2-I (Lozier et al., 1984), the point that the disulfide bonds exist within SCR units in a 1–3, 2–4 pattern was not formally made. This report provides the experimental evidence for C4BP, one of the regulatory proteins of complement, that disulfide bonds may, indeed, be localized within each SCR unit with the first and third and the second and fourth half-cystines in each unit being linked (Figure 5). This finding is consistent with the observation that the SCR units can usually be defined by intron/exon boundaries with each SCR being encoded by a separate exon (Campbell et al., 1984; Leonard et al., 1985; Lintin et al., 1987; Vik et al., 1988). Thus, the SCR unit may represent a basic structural module that may be common to a number of extracytoplasmic proteins (Table I).

The presence of an unusual secondary structure was already inferred from circular dichroism spectra for factor H (DiScipio & Hugli, 1982). It had also been shown that factor H (Sim & DiScipio, 1982) and C4BP (Dählback et al., 1983) are elongated molecules. The elongated shape of these, and presumably of other structurally related proteins (Table I), may be explained by the assembly of the SCR units in tandem fashion, the length of the polypeptide chain being related to the function of a given protein. All these proteins apparently evolved by multiple duplication of an ancesteral SCR unit.

The main structural features seen in the *SCR units*, i.e., the double-loop amino acid sequence generated by the two conserved disulfide bonds [C4BP (Figure 5 of this paper) and β_2-I (Figure 4 of Davie et al. (1986))], show resemblance to structures seen in the *type II domain of fibronectin* (Holland et al., 1987), and also a part of the *kringle* observed in prothrombin (Magnusson et al., 1975), plasminogen (Magnusson

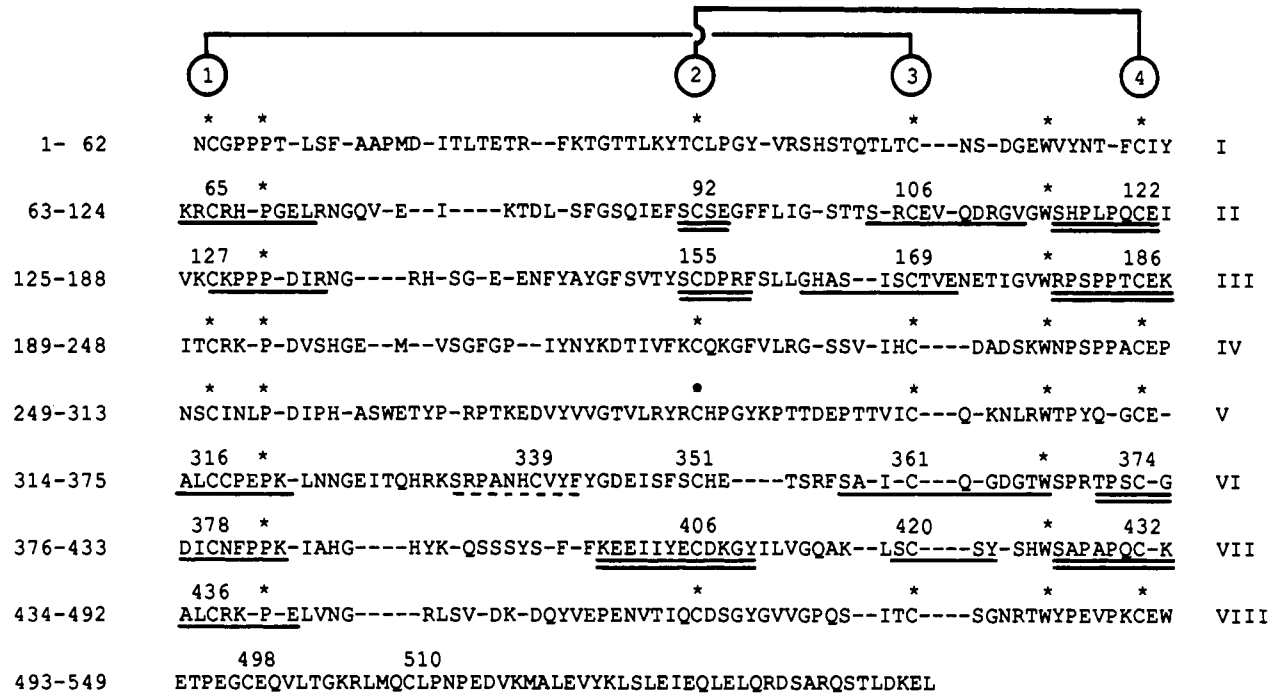

FIGURE 4: Localization of disulfide linkages within the C4BP amino acid sequence. Alignment of the proposed C4BP amino acid sequence (Chung et al., 1985a) to show eight SCR units (I–VIII) and the non-SCR carboxy-terminal region (493–549). The invariant amino acid residues (four Cys, one Pro, and one Trp) were used to align the internal homology regions (asterisks). Gaps have been introduced to improve homology (dashes). The identified pairs of disulfide-bonded peptides establish the first and third (underlined with a single line) and the second and fourth (underlined with a double line) linkages within the SCR units II, III, VI, and VII. A peptide containing Cys_{339} (dashed line) within the SCR unit VI is apparently bound to Cys_{317}. These are two out of four nonconserved half-cystines within the C4BP chain. The other two (Cys_{498} and Cys_{510}) from the C-terminal non-SCR region are presumed to be involved in the interchain linkages. The latter are missing, however, in the C4BP of mouse (Kristensen et al., 1987).

et al., 1976), and some other proteins of the coagulation system (Holland & Blake, 1987). A kringle domain is one of the four different types of structural modules that are present within the regulatory regions of proenzymes from blood coagulation and fibrinolytic systems. The function of these noncatalytic segments is in determining the binding and substrate specificities of plasma serine proteases (Patthy, 1985; Holland & Blake, 1987).

On the comparison of the double-loop sequence of C4BP with the triple-loop sequence of prothrombin fragment 1 (Magnusson et al., 1975), there is a similarity between an SCR unit (58–64 residues long) and a part of the kringle (57 residues long) corresponding to the sequence between residues 84 and 140 of prothrombin fragment 1 (Park & Tulinsky, 1986), with respect to the positions of the two disulfide bridges, and the number of residues between the corresponding cysteinyl residues (Figure 5). The short distance between the midpoints of the two disulfide bridges (~4.1 Å), the two antiparallel strands of highly conserved β-structure, and the stacking of some conserved aromatic residues form the nucleus of the unique chain folding seen in the kringle sequence (Park & Tulinsky, 1986) and type II domain of fibronectin (Holland et al., 1987). As judged by X-ray crystallography, the part of the kringle structure which may be similar in some respects to the SCR unit is folded into three loops. The existence of ω-loops as a novel category of secondary structure has been recognized and defined only recently upon analyses of the three-dimensional structure of globular proteins (Leszczynski & Rose, 1986). It could be speculated that a structural motif of the SCR unit may also consist of three loops centered around the two disulfide bridges. The conserved aromatic residues, one Trp and three Tyr/Phe, within each SCR unit (marked in Figure 5) would be expected to play an important role in the overall folding of each SCR unit. The prothrombin sequences NFCRNPD and GPWCYT around Cys_{115} and Cys_{127}, which are conserved in kringles and form antiparallel β-strands in crystal (Park & Tulinsky, 1986), do not seem to have the equivalent sequences in SCR units of C4BP. However, it follows from the recent computer analyses of a total of 101 SCR sequences that the regions around the second and third cysteinyl residues within each SCR unit are very likely to assume β-structure (Perkins et al., 1988).

In conclusion, all the available data suggest the possibility of similar but not homologous protein fold(s) in the SCR units (Figure 5) when compared to prothrombin kringle 1 (Park & Tulinsky, 1986) or type II domain of fibronectin (Davie et al., 1986; Holland et al., 1987).

Many plasma and membrane proteins are known to consist of the assembly of several types of structural modules. This property is seen in the proenzymes of coagulation and fibrinolytic systems (Holland & Blake, 1987) and also several complement proenzymes, such as C1r, C1s, B, and C2 (Table I). However, the complement regulatory proteins C4BP, H, DAF, MCP, CR1, and CR2 and noncomplement proteins factor XIII and β_2-I seem to consist, with the exception of the C-terminal transmembrane and cytoplasmic regions, of only one type of structural module, the SCR unit. In addition to 101 SCR sequences that were analyzed with the aim to predict the secondary structure of the SCR unit (Perkins et al., 1988), another 56 SCR sequences have been reported [6 in murine C4BP (Kristensen et al., 1987), 4 in DAF (Medof et al., 1987), 4 in MCP (Lublin et al., 1988), 7 in CR1 (Klickstein et al., 1988), 15 in CR2 (Weis et al., 1988), 16 in CR2 (Moore et al., 1987), and 4 in a 35-kDa secretory polypeptide of vaccinia virus (Kotwal & Moss, 1988)]. Thus, the SCR unit seems to be one of the structural modules that is often found encoded by a separate exon, and it is present in a large number of copies. The fact that the genes for CR1, CR2, DAF, MCP, C4BP, H, and factor XIII (Table I) are all located within the human chromosome 1 provides further support for the notion

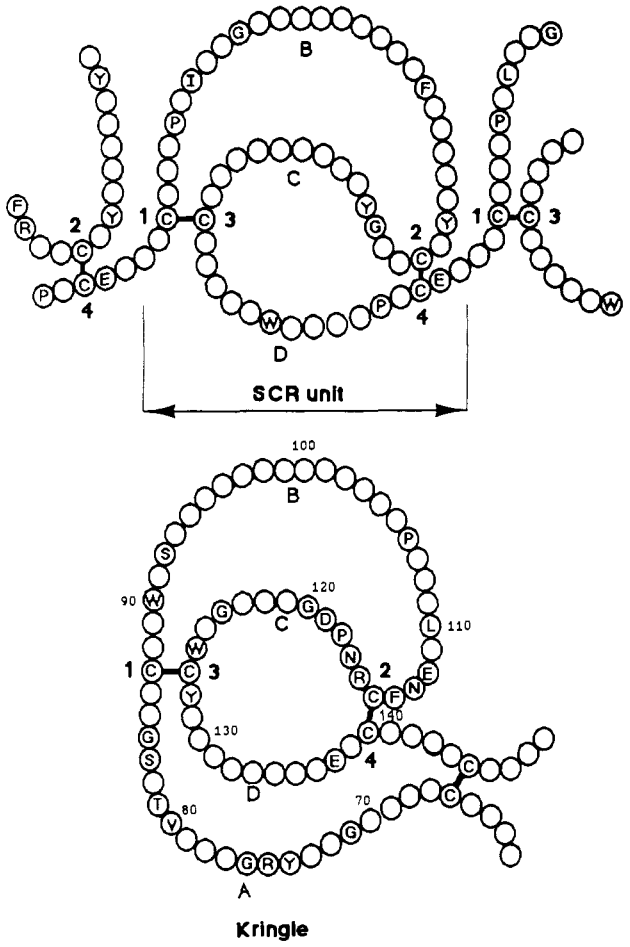

FIGURE 5: Comparison of two structural modules: the SCR unit (C4BP) and the kringle (prothrombin fragment 1). On average, there are 28/29 amino acid residues present between the first and second half-cystines of the SCR unit (in C4BP, numbers range from 26 to 34 residues). This part of the SCR unit corresponds to the B segment (28 residues) of the kringle present in prothrombin fragment 1 (Park & Tulinsky, 1986). The average numbers of 13 residues between the second and third and 12 amino acid residues between the third and fourth half-cystines in the SCR units agree well with the numbers of residues between the corresponding half-cystines in segments C (12 residues) and D (12 residues) of the kringle discussed above. In analogy with the kringle structure of prothrombin fragment 1 (Park & Tulinsky, 1986), it can be assumed that the interaction of two disulfide bridges in each SCR unit, together with the conserved aromatic residues (marked Trp, Tyr, Phe), and several other loop-favoring residues (marked Pro and Gly) may play an important role in the overall three-dimensional folding of SCR unit. See the text for further discussion.

that all these proteins evolved from a common ancestral SCR unit.

ACKNOWLEDGMENTS

We thank Dr. William R. Gray for suggesting use of the two-dimensional reverse-phase HPLC for determination of disulfide-bonding patterns in proteins and for critically reading the manuscript. We are indebted to Dr. Sue K. Holland for fruitful discussion. We also thank Tony Gascoyne for technical assistance and Donald G. Morse for preparation of the figures.

Registry No. C4BP, 87843-67-2.

REFERENCES

Bentley, D. R. (1986) *Biochem. J. 239*, 339–345.

Campbell, R. D., Bentley, D. R., & Morley, B. J. (1984) *Philos. Trans. R. Soc. London, B 306*, 367–378.

Carroll, M. C., Alicot, E. M., Katzman, P. J., Klickstein, L. B., Smith, J. A., & Fearon, D. T. (1988) *J. Exp. Med. 167*, 1271.

Chung, L. P., & Reid, K. B. M. (1985) *Biosci. Rep. 5*, 855–865.

Chung, L. P., Bentley, D. R., & Reid, K. B. M. (1985a) *Biochem. J. 230*, 133–141.

Chung, L. P., Gagnon, J., & Reid, K. B. M. (1985b) *Mol. Immunol. 22*, 427–435.

Dählback, B., & Müller-Eberhard, H. J. (1984) *J. Biol. Chem. 259*, 11631–11634.

Dählback, B., Smith, C. A., & Müller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3461–3465.

Davie, E. W., Ichinose, A., & Leytus, S. P. (1986) *Cold Spring Harbor Symp. Quant. Biol. 51*, 509–514.

Decker, L. A. (1977) *Worthington Enzyme Manual*, Worthington Biochemical Corp., Freehold, NJ.

DiScipio, R. G., & Hugli, T. E. (1982) *Biochim. Biophys. Acta 709*, 58–64.

Drapeau, G. R., Boily, Y., & Houmard, J. (1972) *J. Biol. Chem. 247*, 6720–6726.

Fearon, D. T. (1980) *J. Exp. Med. 152*, 20–30.

Fujita, T., Gigli, I., & Nussenzweig, V. (1978) *J. Exp. Med. 148*, 1044–1051.

Gigli, I., Porter, R. R., & Sim, R. B. (1976) *Biochem. J. 157*, 541–548.

Heinrikson, R. L., & Meredith, S. C. (1984) *Anal. Biochem. 136*, 65–74.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.

Holers, V. M., Cole, J. L., Lublin, D. M., Seya, T., & Atkinson, J. P. (1985) *Immunol. Today 6*, 188–192.

Holland, S. K., & Blake, C. C. F. (1987) *BioSystems 20*, 181–206.

Holland, S. K., Harlos, K., & Blake, C. C. F. (1987) *EMBO J. 6*, 1875–1880.

Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506–3509.

Ichinose, A., McMullen, B. A., Fujikawa, K., & Davie, E. W. (1986) *Biochemistry 25*, 4633–4638.

Janatova, J. (1986) *Biochem. J. 233*, 819–825.

Janatova, J., Reid, K. B. M., & Willis, A. C. (1987) *Complement 4*, 173–174.

Janatova, J., Reid, K. B. M., & Willis, A. C. (1988) *FASEB J. 2*, A1832.

Journet, A., & Tosi, M. (1986) *Biochem. J. 240*, 783–787.

Kinoshita, T., Medof, M. E., & Nussenzweig, V. (1986) *J. Immunol. 136*, 3390–3395.

Klickstein, L. B., Wong, W. W., Smith, J. A., Weis, J. H., Wilson, J. G., & Fearon, D. T. (1987) *J. Exp. Med. 165*, 1095–1112.

Klickstein, L. B., Rabson, L. D., Wong, W. W., Bartow, T. J., Smith, J. A., & Fearon, D. T. (1988) *FASEB J. 2*, A1832.

Kotwal, G. J., & Moss, B. (1988) *Nature 335*, 176–178.

Kristensen, T., & Tack, B. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3963–3967.

Kristensen, T., Ogata, R. T., Chung, L. P., Reid, K. B. M., & Tack, B. F. (1987) *Biochemistry 26*, 4668–4674.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Leonard, W. J., Depper, J. M., Kanehisa, M., Kronke, M., Peffer, N. J., Svetlik, P. B., Sullivan, M., & Greene, W. C. (1985) *Science (Washington, D.C.) 230*, 633–639.

Leszczynski, J. F., & Rose, G. D. (1986) *Science 234*, 849–855.
Leytus, S. P., Kurachi, K., Sakariassen, K. S., & Davie, E. W. (1986) *Biochemistry 25*, 4855–4863.
Lintin, S. J., Lewin, A., & Reid, K. B. M. (1987) *Complement 4*, 186.
Lozier, J., Takahashi, N., & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3640–3644.
Lublin, D. M., Liszewski, M. K., Post, T. W., Arce, M. A., LeBeau, M. M., Rebentisch, M. B., Lemons, R. S., Seya, T., & Atkinson, J. P. (1988) *J. Exp. Med. 168*, 181–194.
Mackinnon, C. M., Carter, P. E., Smyth, S. J., Dunbar, B., & Fothergill, J. E. (1987) *Eur. J. Biochem. 169*, 547–553.
Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123–149, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G., & Clayes, H. (1976) in *Proteolysis and Physiological Regulation* (Robbins, D. W., & Braw, K., Eds.) pp 203–238, Academic Press, New York.
Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199–11203.
Medof, M. E., Iida, K., Mold, C., & Nussenzweig, V. (1982) *J. Exp. Med. 156*, 1739–1754.
Medof, M. E., Lublin, D. M., Holers, V. M., Ayers, D. J., Getty, R. R., Leykam, J. F., Atkinson, J. P., & Tykocinski, M. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2007–2011.
Mole, J. E., Anderson, J. K., Davison, E. A., & Woods, D. E. (1984) *J. Biol. Chem. 259*, 3407–3412.
Moore, M. D., Cooper, N. R., Tack, B. F., & Nemerow, G. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 9194–9198.
Morley, B. J., & Campbell, R. D. (1984) *EMBO J. 3*, 153–157.
Nicholson-Weller, A., Burge, J., Fearon, D. T., Weller, P. F., & Austen, K. F. (1982) *J. Immunol. 129*, 184–189.
Park, C. H., & Tulinsky, A. (1986) *Biochemistry 25*, 3977–3982.
Patthy, L. (1985) *Cell 41*, 657–663.
Perkins, S. J., Chung, L. P., & Reid, K. B. M. (1986) *Biochem. J. 233*, 799–807.
Perkins, S. J., Haris, P. I., Sim, R. B., & Chapman, D. (1988) *Biochemistry 27*, 4004–4012.
Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T., & Tack, B. F. (1986) *Immunol. Today 7*, 230–234.
Rey-Campos, J., Rubinstein, P., & Rodriguez de Cordoba, S. (1988) *J. Exp. Med. 167*, 664–669.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Ripoche, J., Day, A. J., Harris, T. J. R., & Sim, R. B. (1988) *Biochem. J. 249*, 593–602.
Rodriguez de Cordoba, S., Lublin, D. M., Rubinstein, P., & Atkinson, J. P. (1985) *J. Exp. Med. 161*, 1189–1195.
Rodriguez de Cordoba, S., Rey-Campos, J., Dykes, D. D., McAlpine, P. J., Wong, P., & Rubinstein, P. (1988) *Immunogenetics 28*, 452–454.
Sim, R. B., & DiScipio, R. G. (1982) *Biochem. J. 205*, 285–293.
Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261–287.
Tosi, M., Duponchel, C., Meo, T., & Julier, C. (1987) *Biochemistry 26*, 8516–8524.
Vik, D. P., Keeney, J. B., Muñoz-Cânoves, P., Chaplin, D. D., & Tack, B. F. (1988) *J. Biol. Chem. 263*, 16720–16724.
Villiers, M.-B., Reboul, A., Thielens, N. M., & Colomb, M. G. (1981) *FEBS Lett. 132*, 49–54.
Weis, J. J., Toothaker, L. E., Smith, J. A., Weis, J. H., & Fearon, D. T. (1988) *J. Exp. Med. 167*, 1047–1066.
Whaley, K., & Ruddy, S. (1976) *J. Exp. Med. 144*, 1147–1163.
Wong, W. W., Klickstein, L. B., Smith, J. A., Weis, J. H., & Fearon, D. T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7711–7715.