A Study of the Structure of Human Complement Component Factor H by Fourier Transform Infrared Spectroscopy and Secondary Structure Averaging Methods†

Stephen J. Perkins,*,‡ Parvez I. Haris,‡ Robert B. Sim,§ and Dennis Chapman‡

Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K., and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

Received October 5, 1987; Revised Manuscript Received January 8, 1988

ABSTRACT: Fourier transform infrared spectroscopy was used to investigate the secondary structure of human complement component factor H in H_2O and ^2H_2O buffers. The spectra show a broad amide I band which after second-derivative calculations is shown to be composed of three components at 1645, 1663, and 1685 cm^{-1} in H_2O and at 1638, 1661, and 1680 cm^{-1} in ^2H_2O. The frequencies of these components are consistent with the existence of an extensive antiparallel β-strand secondary structure. The exchange properties of the amide protons of factor H as measured in ^2H_2O buffers are rapid and lead to an estimate of NH proton nonexchange that is comparable with those for small globular proteins. Human factor H is constructed from a linear sequence of 20 short consensus repeats with a mean of 61 residues in each one. To investigate the secondary structure further, secondary structure predictions were carried out on the basis of an alignment scheme for 101 sequences for these repeats as found in human factor H and 12 other proteins. These predictions were averaged in order to improve the reliability of the calculations. Both the Robson and the Chou–Fasman methods indicate significant β-structural contents. Residues 21–51 in the 61-residue repeat show a clear prediction of four strands of β-structure and four β-turns. A structural model based on antiparallel β-strands in the secondary structure is proposed and discussed. This is able to account for the pattern of hydrophobic and hydrophilic residues in the repeats, the location of strongly conserved residues and some insertion sites, and the length of 4.5 nm of the 61-residue repeat as determined from previous electron microscopy and synchrotron X-ray scattering studies.

Factor H is a regulatory component of the complement cascade of immune defense and functions as a cofactor for the enzyme factor I in the breakdown of C3b of complement to form iC3b [for recent reviews, see Reid (1983, 1986) and Holers et al. (1985)]. The sequences of mouse and human factor H have been determined from the cDNA (Kristensen & Tack, 1986; Ripoche et al., 1988). Both these proteins are composed entirely of 20 repetitive sequences of polypeptide known as short consensus repeats (SCRs),[1] where each sequence is 61 residues long. These sequences have also been found in other complement components which have the property of binding to C3b or C4b, namely, C2, factor B, C1r, and C1s. The SCRs feature most prominently in other complement regulatory proteins, namely, C4BP and the complement receptors types 1 and 2 and the decay accelerating factor, and have been proposed to constitute a structural superfamily (Reid et al., 1986; Caras et al., 1987). The SCR also occurs in the noncomplement proteins β_2-glycoprotein I, interleukin 2 receptor, factor XIII, and haptoglobin. The three-dimensional structure determination of the SCR would therefore be of great significance in clarifying the nature of the very specific interactions between the complement SCRs and C3b and C4b and the reasons for the occurrence of the SCR in noncomplement proteins.

Factor H is ideal for solution studies of the structural properties of the SCR for reason of its solubility and the simplicity of its chemical structure. Circular dichroism studies have not lead to any inferences of any α-helical or β-sheet conformations (DiScipio & Hugli, 1982). Partial reduction of the protein altered the circular dichroism spectrum to show mainly β-structure and some α-structure. Possibly therefore, and bearing in mind the discussion of systematic errors by Hennessey and Johnson (1982), the relatively high amount of 80 Cys residues (out of a total of 1211) obscures the interpretation of the circular dichroism work. Fourier transform infrared (FTIR) spectroscopy is however able to lead to direct conclusions on the secondary structure in factor H from studies of the fine structure of the amide I band (Susi, 1972; Susi & Byler, 1986; Lee & Chapman, 1986). In FTIR, the use of modern second-derivative methods (Susi & Byler, 1983; Haris et al., 1986) permits the accurate assessment of the number and positions of the component peaks in the amide I band. The application of FTIR spectroscopy to factor H is complemented by comparisons with secondary structure predictions of the main-chain conformation in the SCR. While the accuracy of these methods is variously estimated to lie in a range of 50–60% (Busetta & Hospital, 1982; Kabsch & Sander, 1983a; Nishikawa, 1983), here the predictive method is applied simultaneously to compute the average prediction for all 101 SCR sequences that are currently available in order to improve the precision of the calculations. It is shown from the FTIR

† Financially supported by the Science and Engineering Research Council and the Wellcome Trust. P.I.H. is supported by a SERC CASE award sponsored by Smith Kline & French. This work was presented as a late abstract and poster at the XIIth International Complement Workshop, Sept 18–21, 1987, Chamonix, France.

* Author to whom correspondence should be addressed.
‡ Royal Free Hospital School of Medicine.
§ University of Oxford.

[1] Abbreviations: SCR, short consensus repeat; FTIR, Fourier transform infrared spectroscopy; EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/88/0427-4004$01.50/0

and predictive methods that β-structure can be clearly identified in most of the SCR sequence, and a structural model is proposed and discussed in terms of the properties of factor H.

Materials and Methods

Purification of Human Factor H. Factor H was purified by monoclonal antibody affinity chromatography (B. E. Moffatt and R. B. Sim, in preparation), with fresh citrated human plasma as the starting material. Factor H purified in this way is equivalent in purity, stability, and activity to factor H purified by the conventional chromatography procedures described by Sim and DiScipio (1982). Plasma was made 5 mM with EDTA and then depleted of plasmin/plasminogen by passage over a column of Sepharose–lysine, in the conditions described by Tack and Prahl (1976). The plasma was then passed through a column of Sepharose–horse IgG (10 mg of IgG/mL of Sepharose; 20-mL column for 350 mL of plasma), equilibrated in 25 mM Tris-HCl, 140 mM NaCl, and 5 mM EDTA, pH 7.4. This procedure removes material which interacts with nonimmune aggregated IgG. The plasma was finally passed through a column (10 mL) of Sepharose to which was bound (8.8 mg/mL of Sepharose) the anti-factor H monoclonal antibody MRC OX23 (Sim et al., 1983). The column was washed with 20 column volumes of the starting buffer (as above) and then with 10 column volumes of the same buffer made 0.5 M NaCl. Finally, factor H was eluted with 1.2 column volumes of 3 M $MgCl_2$. This material required no further purification.

Preparation of Concentrated Factor H. For FTIR spectroscopy, 30 mg of factor H (0.9 mg/mL) in 25 mM Tris-HCl, 140 mM NaCl, and 5 mM EDTA was made 15% (w/v) with poly(ethylene glycol) 4000 (Sigma Chemical Co. Ltd.) by addition of the solid, stirred for 40 min at 4 °C, and then centrifuged for 30 min at 15000*g*. The pellet was redissolved by homogenization with a glass rod in 12.5 mM sodium phosphate containing 200 mM NaCl, pH 7.0. The concentration of this preparation was now 14 mg/mL, calculated from the absorption coefficient A_{280}(1cm,1mg/mL) of 1.42 (Sim & DiScipio, 1982). A further sample was concentrated in the same way to 36 mg/mL. This more concentrated solution was stable at room temperature, but the protein precipitated reversibly on being cooled to 4 °C. This procedure leads to a small contamination of the samples with poly(ethylene glycol) wich may be present at a maximum concentration of 1.4 or 3.6 mg/mL in the concentrated protein samples. Deuterium exchange experiments were performed by dialyzing the factor H solutions against several changes of the same buffer prepared in ^2H_2O at room temperature for 2 days and thereafter for 6 days at 4 °C (14 mg/mL) or 20 °C (36 mg/mL).

FTIR Spectroscopy. Infrared spectra were recorded on a Perkin-Elmer 1750 Fourier transform infrared spectrometer equipped with a TGS detector. A Perkin-Elmer Model 7300 data station was used for data acquisition, storage, and analysis. Samples were placed in a thermostated Beckman FH-01 CFT microcell fitted with CaF_2 windows and a 6-μm tin spacer. The protein samples and its buffer were measured with identical scanning parameters after equilibration at 20 °C for 15 min. Figures 1 and 2 correspond to the difference spectra recorded between the sample and buffer spectra, calculated as in Haris et al. (1986). The sample compartment was continuously purged with dry air to eliminate absorption by water vapor in the spectral region of interest. Poly(ethylene glycol) has no infrared absorption bands between 1450 and 1900 cm^{-1}, which amply avoids the region of interest between 1500 and 1700 cm^{-1} (Pouchert, 1981). The spectra recorded in H_2O buffers (800 scans in 119 min) were signal-averaged at a resolution of 4 cm^{-1}, while those measured in ^2H_2O buffers (200 scans in 55 min) were averaged at a resolution of 2 cm^{-1}. Other details of the methods used for spectral subtraction and for obtaining second-derivative spectra are described in Haris et al. (1986). In order to eliminate noise, the second derivative (rate of change of slope) was calculated over a 13 data point range (13 cm^{-1}).

Secondary Structure Predictions. The complete amino acid sequences of 12 proteins were used as follows: α-chain of human haptoglobin (Kurosky et al., 1980); human and mouse interleukin 2 receptor (Miller et al., 1985; Leonard et al., 1985); human plasma β_2-glycoprotein I (Lozier et al., 1984); b subunit of human factor XIII (Ichinose et al., 1986); human C1r (Leytus et al., 1986; Journet & Tosi, 1986; Arlaud & Gagnon, 1983; Arlaud et al., 1987); human C1s (Mackinnon et al., 1987); human C2 (Bentley, 1986); human factor B (Morley & Campbell, 1984); human C4BP (Chung et al., 1985); human and mouse factor H (Ripoche et al., 1988; Kristensen & Tack, 1986). A partial sequence for human CR1 encompassing 23 SCRs of a total of 30 was used (Klickstein et al., 1987). Between them, these 13 proteins contain a total of 101 SCRs.

Secondary structure predictions were performed by the methods of Robson and Chou–Fasman (Garnier et al., 1976; Chou & Fasman, 1974a,b, 1977, 1978). The programs ROBSON and CHOU1 as employed here were implemented on the VAX 11-750 minicomputer system of Birkbeck College, University of London. For each residue, ROBSON sums up terms extending ± 8 residues from the residue, which will specify the prospensity of these neighboring residues to influence the secondary structure of the residue in question. These terms correspond to each of the four states, extended, helix, turn, and coil. After summation of all the terms over all residues, the highest scoring of the four states for each residue determines which state it is in (Figure 4). Decision constants can be used as cut-offs to bias the predictions in favor of β-structure (Figure 4) or α-helices. CHOU1 assesses short sequences for α-helix and β-sheet nucleation sites on the basis of standard probabilities P_a, P_b, and P_t for each residue. Once these nucleation sites are identified, the secondary structure is extended until structural breakers occur. Conflicts between overlapping α, β, and turn regions can be resolved on the basis of comparing the P_a, P_b, and P_t values for the residues in question, although in this study this was not done (Figure 5). Additional programs PREDRB and PREDCF were written to edit the output from ROBSON and CHOU1: (a) to align the sequences of the 101 SCRs and their structural predictions; (b) to sum up the predictions and provide an analysis of this; (c) to provide statistical and graphical breakdowns of the frequencies of the residue type and their chemical properties and structural analysis. For Chou–Fasman analyses, the structural classes were summed with unit weights if classified as definite and with half-weights if classified as probable or possible.

Results and Discussion

Infrared Spectroscopy of Factor H. Difference infrared spectra of factor H dissolved in H_2O and ^2H_2O buffers are shown in Figure 1. The spectra in ^2H_2O were recorded after 3 h of exchange at room temperature. In H_2O, the difference spectrum shows the maxima of the amide I and amide II bands to occur at 1647 and 1552 cm^{-1}, respectively (Figure 1). In ^2H_2O buffers, there is a dramatic reduction in the intensity of the amide II band at 1552 cm^{-1}, which shifts to approximately 1456 cm^{-1} (not shown). These changes are attributed

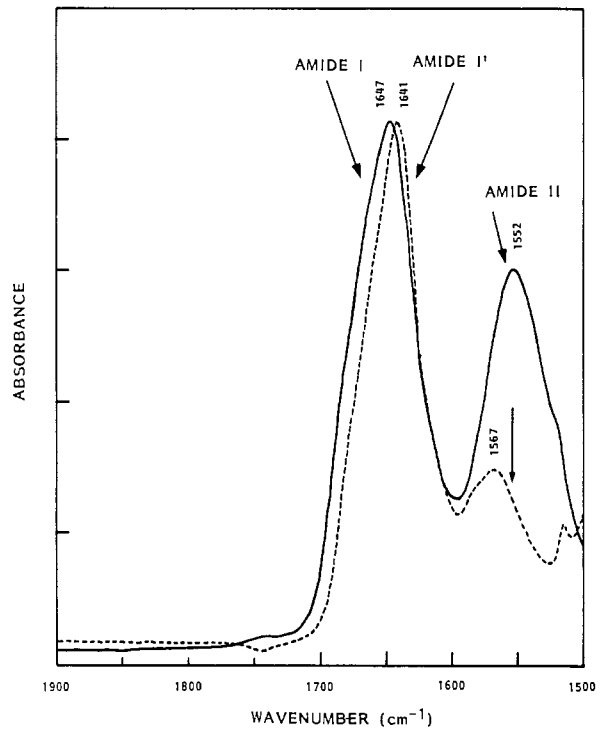

FIGURE 1: FTIR spectra of factor H at 20 °C dissolved in H_2O and ^2H_2O buffers. The protein concentration is 36 mg/mL in buffers of 12.5 mM sodium phosphate, 200 mM NaCl, and 0.5 mM EDTA. The intensity of the amide II band at 1552 cm^{-1} is much diminished as arrowed after dialysis in ^2H_2O buffers for 3 h at room temperature.

to the isotopic substitution of the exchangeable NH protons of the polypeptide by ^2H. There is also a relatively minor shift of the amide I band from 1647 to 1641 cm^{-1} on exchange.

Calculations of the second-derivative spectra were employed in order to resolve the overlapping bands within the amide I band (Susi & Byler, 1983; Haris et al., 1986). Absorption bands in the original difference spectrum (Figure 1) are revealed as negative bands in the second-derivative spectrum (Figure 2). The results of Figure 2 show that the amide I band of factor H has a relatively simple three-component substructure which is independent of concentration effects within a precision of 2 cm^{-1}. The individual components are observed at 1645, 1660–1667, and 1684–1686 cm^{-1} in H_2O and at 1638, 1660–1661, and 1679–1680 cm^{-1} in ^2H_2O in that order. This spectral simplicity is reminiscent of the three-component substructure of the amide I band of hemoglobin (Susi & Byler, 1983), whose secondary structure is α-helical (Kabsch & Sander, 1983b). This is in marked distinction to the five- to six-component substructure of ribonuclease (Olinger et al., 1986; Haris et al., 1986) and the six-component substructure of β-lactoglobulin A (Susi & Byler, 1983), whose structures are composed of α-helices and β-sheets in various ratios.

The use of H–^2H exchange is a powerful tool in the identification of secondary structure types (Olinger et al., 1986; Haris et al., 1986). There is good evidence that the presence of the 1645- and 1685-cm^{-1} bands (H_2O) and the 1638- and 1680-cm^{-1} bands (^2H_2O) indicates the presence of antiparallel β-sheet secondary structure (Susi & Byler, 1986). Indeed, the similar magnitudes of the frequency shifts for both bands (5–7 cm^{-1}) on H–^2H exchange suggest that both bands originate from the same type of secondary structure. The 1645-cm^{-1} (H_2O) and 1638-cm^{-1} (^2H_2O) bands are too low in frequency to be assigned to α-helices (1650–1655 cm^{-1}) but

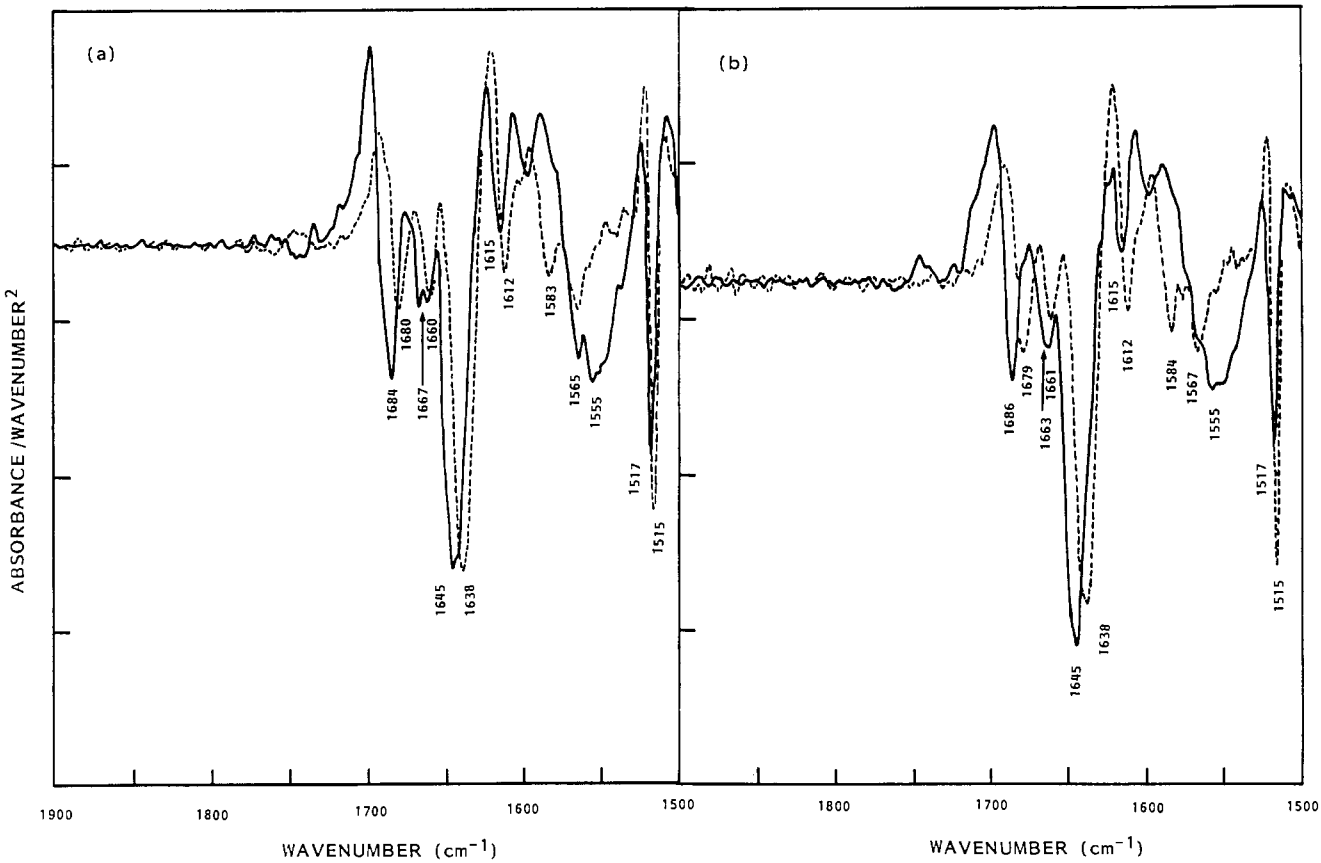

FIGURE 2: Second-derivative FTIR spectra of factor H in H_2O and ^2H_2O buffers at two concentrations. (a) The spectra correspond to those in Figure 1 at 36 mg/mL. (b) Spectra are recorded at 14 mg/mL. The peaks shift as a result of isotopic substitution of NH by N^2H. (—) H_2O; (– –) ^2H_2O.

are fully consistent with the strong low-frequency band of a β-strand (Byler & Susi, 1986; Olinger et al., 1986; Haris et al., 1986). The 1684–1686-cm^{-1} (H_2O) and 1679–1680-cm^{-1} (^2H_2O) bands are assigned to the high-frequency component of this β-strand (Olinger et al., 1986; Haris et al., 1986). The presence of this weak band indicates that the β-strands are antiparallel (Susi et al., 1967; Susi, 1969).

The third weak band is observed at 1660–1667 cm^{-1} in H_2O, and this only shifts by 2 cm^{-1} to 1661 cm^{-1} in ^2H_2O. This band is split in the spectrum run at 36 mg/mL in H_2O but not at 14 mg/mL in H_2O nor in ^2H_2O, and this splitting is considered to be artifactual. The small magnitude of the band shift on H–^{2}H exchange is attributed to either a strongly hydrogen bonded environment or a highly buried one in which it is protected from isotopic exchange. In proteins, turns and bends have been reported to be associated with a band near 1663 cm^{-1} (Byler & Susi, 1986). Given the high amount of β-sheet structure in factor H (above), it is possible that a significant amount of strongly hydrogen bonded β-turns is present in order to mediate the interconnections of the β-strands within the sheet structure, and a consideration of the presence of several semiconserved Gly residues in the sequence (see below) supports this contention. It is therefore reasonable to presume that the 1661–1662-cm^{-1} band is evidence of the existence of β-turns. There is however some ambiguity regarding the precise assignment of this band, since turns, α-helices, and even disordered structures have been suggested to absorb in this region (Olinger et al., 1986; Haris et al., 1986; Byler & Susi, 1986). The absence of other bands corresponding to this structural type may suggest that the presumed turn structure is found in essentially one structural form. Overall it is concluded that factor H has a predominantly antiparallel β-sheet structure and that no α-helix structure is detectable.

In FTIR studies on immunoglobulin G (which has a predominantly β-sheet structure), four bands at 1638, 1663, 1674, and 1690 cm^{-1} are observed in H_2O–saline buffers (Wasacz et al., 1987). The latter two bands were assigned to either a β-sheet or a turn. Since the first two bands are identical with those seen in Factor H and the 1674-cm^{-1} band is not seen in factor H, this suggests that in fact the 1674-cm^{-1} band is assignable to a turn structure and that the 1690-cm^{-1} band results from the β-sheet.

The decrease in intensity (Figure 1) of the amide II band observed at 1552 cm^{-1} in ^2H_2O can be used as a measure of H–^{2}H exchange in peptide groups. Detailed comparisons of second-derivative FTIR spectra of factor H in ^2H_2O after dialysis for 3 h (Figures 1 and 2) and after 24 h at 20 °C show that these are very similar. The further exposure of factor H for another 8 days in ^2H_2O buffer shows no further intensity changes in the amide II band or shifts in the amide I band. This indicates that at p^{2}H 7.3 the majority of structural changes involved in H–^{2}H exchange take place rapidly. Comparison of the spectral intensities at 1552 cm^{-1} before and after exchange suggests an upper limit of 23% of nonexchange of the main-chain amide protons (as arrowed in Figure 1), although the precision of this estimate is affected by the presence of the neighboring, partially overlapping band at 1567 cm^{-1} that is observed in ^2H_2O solution and also by the inability to observe the original signal itself in ^2H_2O. This figure is comparable to small globular proteins such as lysozyme or ribonuclease, where 10–20% of the main-chain NH protons remain nonexchanged after extensive dialysis in ^2H_2O buffers (Perkins, 1986). These data, together with the observation of two large band shifts on dissolution into ^2H_2O solvent, show that the β-sheet structure of factor H is readily accessible to solvent, despite its high M_r of 155 000. This is fully compatible with the highly elongated structure that is expected for factor H (Sim & DiScipio, 1982) by analogy with that for C4BP (Dahlback et al., 1983; Perkins et al., 1986).

Table I: Survey of the Frequency of Residues in the Short Consensus Repeat[a]

segment (residues)	position	residue	frequency
A (1–7)	2	Cys	100
	5	Pro	74
	6	Pro	44
B (8–20)	8	Ile, Leu, Val	37 + 19 + 14 = 70
	10	Asn	61
	11	Gly, Ala, Ser	67 + 14 + 6 = 87
	13	Ile, Leu, Val	21 + 9 + 19 = 49
C (21–39)	21	Phe, Tyr	27 + 49 = 76
	24	Gly	69
	25	Asp, Glu	22 + 24 = 46
	27	Ile, Leu, Val	22 + 17 + 46 = 85
	29	Phe, Tyr	27 + 62 = 89
	31	Cys	100
	34	Gly	67
	35	Phe, Tyr	32 + 49 = 81
	39	Gly	55
D (40–46)	43	Ile, Leu, Val	32 + 5 + 15 = 52
	45	Cys	99
E (47–55)	49	Gly, Ala, Ser	76 + 3 + 8 = 87
	51	Trp	95
F (56–61)	56	Pro	80
	58	Cys	100

[a] If a residue or residue group is present more than 40 times in the 101 sequences, it is noted. The five charged residues occur more than 40 times at the following positions: 9, 12, 25, 26, 28, 30, 32, 59, 60. Residues are shown as the following groups when significant totals of each are found at that position: tiny, G, A, S; aliphatics, I, L, V; aromatics, H, F, Y, W; positive, R, H, K; negative, D, E.

Secondary Structure Prediction of the SCR. Secondary structures were predicted from the amino acid sequences by the Robson method (Garnier et al., 1978). This has advantages: (a) it is slightly more accurate than other commonly used methods such as the Chou–Fasman and the Lim procedures (Lim, 1974a,b; Kabsch & Sander, 1983a; Taylor & Thornton, 1984); (b) since the conformational prediction is objectively chosen as one of four states, this facilitates further computer analyses of the results; (c) it is possible to choose a conformational preference by the use of decision constants which can be set at cut-offs that favor predominantly α-helical or predominantly β-sheet predictions, or a mixture of these. Prior to this study, preliminary calculations have suggested the presence of β-structures in the SCRs of C4BP and the CR1 receptor (Chung et al., 1985; Klickstein et al., 1987).

The SCR occurs 20 times in human factor H and a further 81 times in 12 other proteins whose sequences have been reported (Materials and Methods). The 101 sequences of the SCR were aligned on the basis of maximizing the occurrence of identical or chemically homologous residues (defined in Table I) at a given residue position and minimizing the occurrence of gaps in this alignment. This procedure permits the alignment of the secondary structure predictions for calculation of the average. The alignments of Figure 3 thus differ from others in the original sequence studies, where many more gaps had been included in order to maximize possible residue homologies. The maximum length of an SCR was found to be 98 residues. Since 61 of these residue positons are occupied in over half the sequences examined, this was deemed to be the standard length of an SCR, and residues are accordingly numbered from 1 to 61 (Figure 3). Table I shows that 23 of the 61 positions are conservatively occupied by amino acid residues. In particular, Cys-2, Cys-31, Cys-45, Trp-51, and

FIGURE 3: Alignment of the sequences of 101 short consensus repeats found in 13 proteins. The six conserved segments found after sequence alignment are denoted as A–F, and the residues within these are numbered 1–61. Residues are included in the segments A–F when more than half the positions in the 101 sequences are occupied. The maximum length of a SCR is 98 residues (ignoring the short N-terminal extensions).

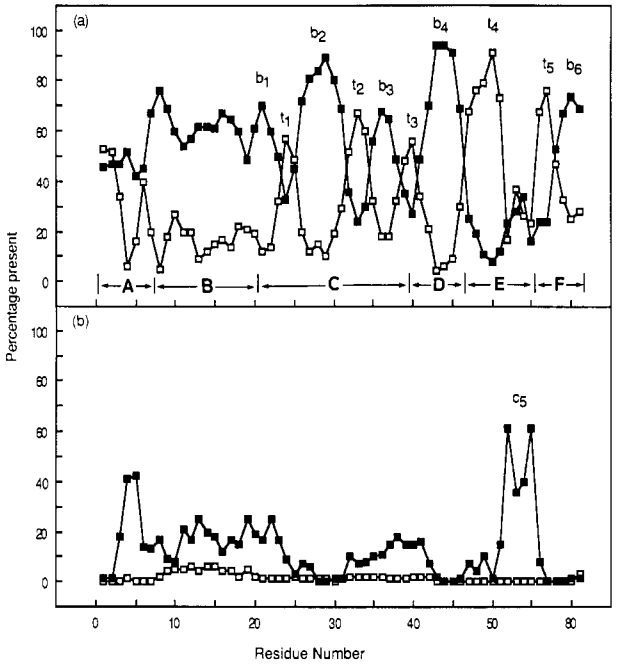

FIGURE 4: Averaged Robson secondary structure prediction for the 61 residues of the short consensus repeats. This is biased toward the prediction of β-structures. (a) The total of predicted extended (■) and turn (□) conformations is plotted against residue position. Note the regular alternation of β-sheet and turn predictions between residues 20 and 52, denoted b_{1-4} and t_{1-4}. The six segments in the SCR are labeled A–F to follow Figure 3. (b) The remaining predictions for helix (□) and coil (■) conformations in the SCR are shown. Note that for a given residue the sum of the four predictions totals 100%.

Cys-58 occur 95 times or more; note that this includes all four Cys residues present in the SCR. A further eight positions (totaling 31 conservative sites in all) are predominantly occupied by charged residues.

The examination of the alignment of Figure 3 shows that six segments of continuous residues could be identified, and these are lettered A–F (Table I). With the exception (ignored) of the residues at the N-terminus of each protein, six insertions were identified and labeled as u–z. Their lengths vary between 0 and 11 residues, depending on the SCR in question.

Since the FTIR spectra of factor H indicated the preponderance of β-sheet structures, Robson predictions were performed with decision constants biased for β-structures, i.e., with cut-offs of 158 and −88 for the α-helix and β-sheet states to correspond to less than 20% and over 20% of occurrences, respectively. The total of each of the four conformational states in the 101 sequence predictions for the 61 SCR residue positions is summarized in Figure 4. Since the accuracy of the Robson method is only about 55% for a given residue (Kabsch & Sander, 1983a), which is to be compared with an accuracy of 25% had a method of random predictions been employed instead, the summation of Figure 4 is expected to highlight the structural predictions which are significantly close to the actual secondary structure. The effect of any minor errors in the sequence alignment of Figure 3 will also be minimized by the use of the summation method. Thus inspection of Figure 4 shows in confirmation of these expectations that many residues are neatly predicted to occur as one structure in 50–70% of the predictions and are readily distinguished from the other three structures which occur at levels of about 0–25%. The use of unbiased and α-helix-biased decision constants does not alter the overall appearance of well-defined maxima and minima, although the levels of the predicted percentages of residues in the α- and β-states move up and down, respectively. The use of reduced totals of SCRs increases the level of statistical noise, while the general pattern of peaks is reproduced, as to be expected.

The most notable features of the structural predictions are summarized as follows:

(a) β-Sheet structures are predicted to be the dominant conformation in 41 of the 61 positions (Figure 4a). Their occurrence is also common with the use of unbiased decision constants. Segments A and B are predominantly β-sheet. Segment C is of particular interest in that three segments of β-strands are predicted (b_1, b_2, and b_3 in Figure 4a), between which two sharp well-defined turns are found (t_1 and t_2). Segment D corresponds to another β-strand prediction, b_4, which is flanked by two turn predictions, t_3 and t_4. Segment F corresponds to a short β-strand, b_6, which is preceded by another turn prediction, t_5. These occurrences are further discussed below.

(b) The proportion of α-helix predictions is very low in Figure 4b. Even when the decision constants are set to force the prediction of α-helices, such α-helices do not dominate the predictions as the β-strands and β-turns do in Figure 4a. While segment B could constitute a complete α-helix in this extreme prediction, this structure was only slightly ahead of the totals of turn and coil predictions.

(c) After the β-strand, turn predictions are the next most dominant form and occur in 16 positions. The most prominent occurrences are located at positions 24 and 25 (denoted t_1), 32, 33, and 34 (t_2), 39 and 40 (t_3), 47, 48, 49, 50, and 51 (t_4), and 56 and 57 (t_5). With the exception of turn t_4, most of these predictions correspond to sharp peaks. In the case of turn t_4, it should be noted from Figure 3 that 29 and 22 gaps in the 101 sequences were introduced at positions 48 and 50 in order to retain Gly-49 as a conserved residue, which lengthens the turn. Other weaker turn occurrences are found at position 5 (which is a conserved Pro site), positions 9–11, and position 17.

(d) The coil state is predicted for the four contiguous positions 52, 53, 54, and 55 (Figure 4b) and is a sharp feature in an otherwise generally featureless plot. Comparison with Figure 4a shows that this segment c_5 occurs neatly between the turns t_4 and t_5. A weaker coil prediction is located at residues 4 and 5.

In conclusion, the analyses of the summed predictions in Figure 4 show evidence for β-strand and turn structures in 57 of the 61 residues of the SCR. Interestingly, segments C–F (residues 21–61) constitute a series of structural alterations in which five β-strands and one coil structure are systematically interpolated by five turns. Residues 1–20 are largely predicted as a β-strand, together with some weak evidence for three turns in it. These structural predictions are fully supportive of the FTIR results for human factor H and extend the FTIR prediction of β-structure to the 61 residues of the SCR by locating the likely regions in which it occurs.

Chou–Fasman predictions were carried out as a control of the Robson predictions. Since no choices were made between the predictions of α-helices, β-sheets, and turns, the predicted states overlap. In this sense, the summary of Figure 5 can be considered to be an unbiased structure prediction. For these reasons, the helix prediction of Figure 5b is qualitatively different from that of Figure 4b. In comparison with Figure 4a, Figure 5a shows that the major features b_{1-4} and t_{1-5} are reproducibly well predicted, as is also the minor feature t_5–b_6 at the C-terminus. For residues 1–11, β-sheets now occur at a lower occupancy than that for turns. This is similar to the results of Robson calculations using unbiased decision constants and suggests that the prediction of β-structure in this

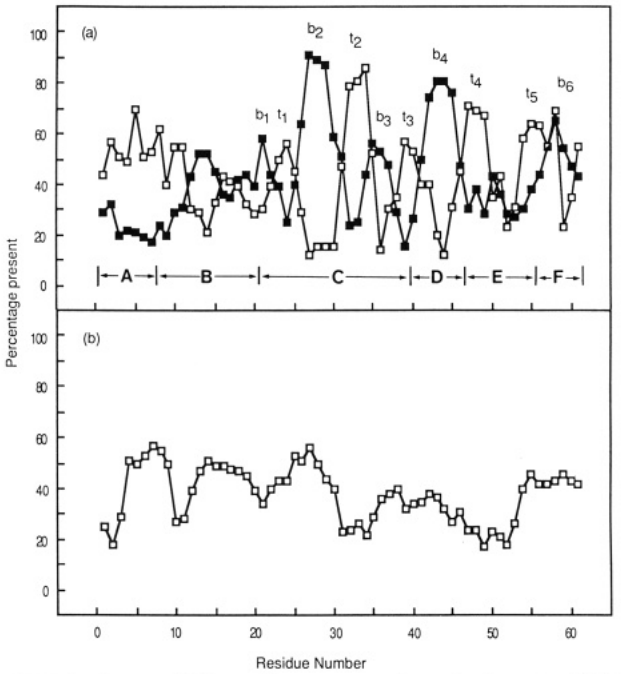

FIGURE 5: Averaged Chou–Fasman secondary structure prediction for the 61 residues of the short consensus repeat. (a) The total percentage of occurrences of β-strand (■) and turns (□) is plotted against residue position. The peaks denoted b_{1-4} and t_{1-4} correspond well with those in Figure 4a. (b) The total of occurrences of predicted α-helices is shown. Note that for a given residue no choice is forced between the α-helix, β-strand, and turn prediction. The vertical axis thus corresponds to the total percentage of each prediction as found in the 101 SCRs.

region is not as strong as for residues 20–61. In addition, the predicted occurrence of α-helices by the two methods is found to be in poor agreement, and this is interpreted as further evidence in support of a general β-sheet structure. For example, only the C-terminal part of segment B is predicted as an α-helix by the Chou–Fasman method (Figure 5a), which is in conflict with the Robson prediction (when biased for α-helices) that suggested that the whole of segment B forms an α-helix. Overall, these comparisons reinforce the conclusions drawn from the analyses of Figure 4.

A Secondary Structure Model for the SCR. The proposed β-structure for the SCR is now discussed in terms of a possible model which will bring together the FTIR results with the secondary structure predictions (Figure 6). The best-defined β-sheet prediction of Figures 4 and 5 corresponds to the three peaks b_1, b_2, and b_3 and the two turns t_1 and t_2 in residues 21–39 of segment C. These are schematically depicted in Figure 6 as a β-sheet with two 3-turns; this is antiparallel in accordance with the FTIR results. Note that secondary structure prediction methods are not able to identify supersecondary features. The b_4 peak which is contained in segment D (residues 40–46) and the broad t_4 turn of the first part of segment E (residues 47–51) are also attached as the continuation of this β-sheet.

The relatively small size of the SCR and the knowledge that it is extensively exposed to solvent (Dahlback et al., 1983; Perkins et al., 1986) imply that one surface of the β-sheet will be solvent exposed. In this case, it is expected that all the odd-numbered residues (say) will have a hydrophobic character, while all the even-numbered residues will have a hydrophilic character, since it is well-known that the relative dispositions of the amino acid side-chain alternate on the two sides of the β-sheet (Lim, 1974a; Cid et al., 1982; Argos & Mohanarao, 1986). The residues of the SCR were examined

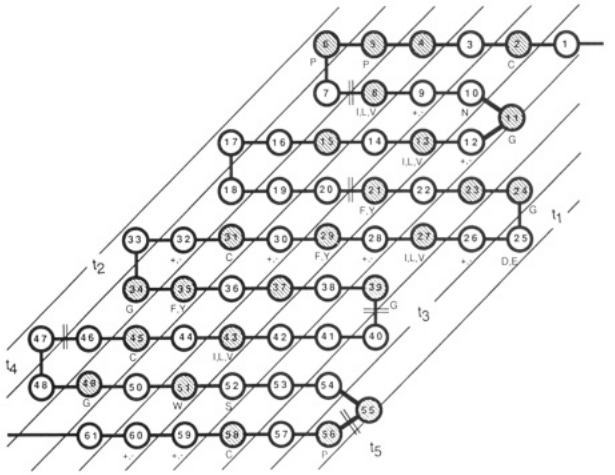

FIGURE 6: Schematic view of a possible antiparallel β-sheet model for the short consensus repeat. Residues 21–51 are well explained in terms of this model, residues 7–20 and 52–61 are reasonably compatible with it, and residues 1–6 are not well explained (see text). Hydrophobic residues are denoted by shaded circles and hydrophilic ones by open circles. Conserved residues noted in Table I are indicated by the single-letter code, or by (+, −) if in the charged group. Double slashes are indicated between the residues which correspond to sites of the residue insertions u–z (see Figure 3). The diagonal lines mark out the bands of alternating hydrophobic and hydrophilic amino acid residues. This alternating pattern is consistent with the possible formation of a β-sheet sandwich with the hydrophobic residues at the interface between the two β-sheets.

for patterns of hydrophobicity and hydrophilicity. Residues were classified with the consensus hydrophobicity scale for residues (Eisenberg, 1984), which is subdivided into the two groups as in Perkins (1986). In Figure 6, the summations over the 101 SCR sequences generally show as expected a clear alternation in the occurrence of hydrophobic and hydrophilic residues between residues 21–51. There are six exceptions. Gly-24 and Gly-34 correspond to hydrophobic residues in what are anticipated to be hydrophilic sites. However, these residues have the smallest side chains and occur frequently at turns (Chou & Fasman, 1977, 1978), and these are compatible with the model of Figure 6. The other four exceptions at positions 25, 33, 41, and 47 are hydrophilic residues which occupy putative hydrophobic sites. All of these are located at turns, where it is generally recognized that such residues are usually hydrophilic (Rose, 1978; Chou & Fasman, 1978; Schulz & Schirmer, 1979). In conclusion, residues 21–39 are fully consistent with the hydrophobic and hydrophilic properties to be expected from a β-sheet structure, one surface of which is exposed to water.

The antiparallel β-sheet model for residues 21–51 is supported further by the positions of the two Cys residues Cys-31 and Cys-45. Both are on the hydrophobic side of the β-structure, which is to be expected if the internal disulfide bonds are to be formed between Cys-31 and Cys-58 and between Cys-45 and Cys-2 (Lozier et al., 1984; Day et al., 1987; Janatova et al., 1987). It is also noteworthy that the strands b_1–b_4 are up to nine residues in length; the typical β-strand or β-ladder is no more than nine residues long (Kabsch & Sander, 1983b; Cohen et al., 1986). It is also noted that the insertion segment x occurs between residues 39 and 40 and segment y between residues 46 and 47. Both sides are coincident with the surface-exposed turns t_3 and t_4. It can be readily appreciated how these insertions can be attached as surface loops on the SCR structure.

A reasonably satisfactory correlation of residues 52–61 with an antiparallel β-sheet can be proposed. For the purpose of discussion, these residues as well as those of residues 1–20 are

also incorporated into the schematic diagram of Figure 6, even though the structural model for these residues is not as strong as that for residues 21–51. Residues 52–61 demonstrate the requisite pattern of alternating hydrophobicity and hydrophilicity with only two exceptions if residue 55 and the insertion z are postulated to occur at a 4-turn (Kabsch & Sander, 1983b; Milner-White & Poet, 1986). The scheme of Figure 6 is interesting in that Cys-58 has the same horizontal coordinate as Cys-31. This is suggestive of how the disulfide link can be formed if the two-dimensional representation is postulated to fold into a three-dimensional β-sheet sandwich of the orthogonal or aligned types discussed in Chothia and Janin (1982). That residue 53 is not hydrophobic is attributed to the prediction of the coil state for residues 52–55 and its proximity to the proposed turn t_5. That residue 60 is not hydrophobic can be accounted for in terms of its location at the edge of the β-sheet (Garrett et al., 1985).

The β-sheet model for 1–20 is not satisfactory. The pattern of hydrophobicity and hydrophilicity is still observed if Gly-11 is located at a 4-turn as shown. The possible turns at residues 9–13 and 16–19 bear a hydrophilic character as desired; however, neither of these is strongly predicted in Figure 4 or 5. The situation with regard to residues 1–6 is unclear. The averaged predictions are also weakly suggestive of a further turn at residues 6–7 in the simple model of Figure 6. In this case, the N-terminus will lie at the opposite end of the β-sheet to the C-terminus. Such an arrangement in Figure 6 places Cys-2 at a distance from Cys-45 in the horizontal direction, which is not compatible with the requirement of disulfide bridge formation in the possible β-sheet sandwich model. This can be overcome by redeploying residues 1–6 to the left from the location shown in Figure 6, whereupon the horizontal coordinates of Cys-2 and Cys-45 become similar.

Conclusions

From FTIR spectroscopy on human factor H, it is concluded that the short consensus repeat has a structure that is typical of small globular proteins and is found in a predominantly antiparallel β-sheet conformation. The use of summation of the 101 individual secondary structure predictions to obtain an average has been of value in this study in that only the significantly predicted features are retained for further structural analyses. These have suggested the likely existence of a structure with extensive β-strands and turns in the short consensus repeat. This averaging constitutes a simple method that may generally improve the reliability of secondary structure predictive schemes beyond the current estimates of 50–60% accuracy. Combination of the FTIR and predictive results in the schematic model of Figure 6 shows that residues 21–51 can be well explained in terms of an antiparallel β-sheet and that residues 7–20 and 52–61 are reasonably compatible with the β-sheet model, while the situation in relation to residues 1–6 is less clear. The combination of FTIR spectral interpretation with predictions of protein secondary structures appears to be a promising technique. Further assessment of this combined approach using a large set of proteins of known crystallographic structures should indicate the degree of validity of this technique.

Reasons for the occurrence of conservative or semiconservative amino acid residues in the SCR can be proposed or commented upon. In physiological terms, the most striking aspect of this model is the arrangement of five semiconserved sites for charged residues at positions 25, 26, 28, 30, and 32 (Table I). All of these are located on strand b_2 (Figure 6), which is one of the best predicted strands in the analyses of Figures 4a and 5a. It is quite possible that variations in the nature of these charged groups may play an important role in determining the biological specificity of each SCR within the parent protein structure in terms of the highly specific long-range protein–protein interactions that take place in serum between factor H and C3b, C4BP and C4b, and so on. Another set of semiconservative residues occurs 67, 69, 67, 55, and 76 times out of 101 at Gly-11, Gly-24, Gly-34, Gly-39, and Gly-49 (Table I). These are frequently found at turns in proteins, and the observed positions in Figure 6 show that this is the likely explanation for all the five semiconserved Gly positions. In a type II reverse turn, an obligatory Gly residue is located at position $i + 2$ (Schulz & Schirmer, 1979). This is the case for the turn t_2 but not for the turns t_1, t_3, and t_4, although it should be added that other physical techniques such as two-dimensional NMR or protein crystallography are expected to refine significantly the simple scheme depicted in Figure 6. The relatively frequent occurrences of seven Ile,-Leu,Val and Phe,Tyr sites (Table I) is of great interest since these residues are known to occur at significantly higher proportions in β-sheet to β-sheet interfaces and serve to make the internal surface of the two β-sheets approximately smooth (Chothia & Janin, 1981, 1982). Together with the highly conserved Cys residues at Cys-2, Cys-31, Cys-45, and Cys-58, which are important for defining the relative configuration of the two β-sheets, these considerations indicate straightforward reasons for the occurrence of most of the residues noted in Table I.

Some three-dimensional structural information on the SCR is available for C4BP from electron microscopy, solution scattering, and hydrodynamic analyses (Dahlback et al., 1983; Perkins et al., 1986). From these, it is deduced that a length of 29 nm corresponds to 6.5 SCR sequences; thus the SCRs are separated by approximately 4.5 nm between their centers. From the sequences of mouse and human factor H, crystallographic volumes of 173.0–173.7 nm^3 were calculated for the protein component only (Chothia, 1975; Perkins, 1986). The volume of a single ellipsoid defining the shape of the SCR is therefore 8.68 nm^3. These data can be compared with the possible β-sheet model of Figure 6. From Figure 6, the overall length of the β-sheet model is 11–12 residues. A mean distance of 0.38 nm between the α-carbon coordinates of a given β-sheet in an extended configuration in the crystal structure of the IgG fragment Fab New is readily computed. The total length of the SCR model in Figure 6 is thus 4.2–4.6 nm. The separation of the two sheets in a β-sheet sandwich lies between 0.83 and 1.13 nm depending on whether their packing is aligned or orthogonal (Cohen et al., 1981; Chothia & Janin, 1982). This implies that the thickness of a β-sheet sandwich is between 1.6 and 2.2 nm. The interchain spacing between two parallel or antiparallel β-strands is 0.485 nm (Ashida et al., 1981). The width of a four β-strand model is therefore about 2 nm. The volume of the ellipsoid of axes 4.5 nm × 1.9 nm × 2.0 nm corresponding to this β-sheet model is 9.0 nm^3, which is in good accord with the value calculated above from sequence data. The overall dimensions of this model are also in good agreement with the center-to-center separation between neighboring SCR domains in C4BP. These dimensions will be compared with models for factor H that are currently being developed from synchrotron X-ray and high-flux neutron scattering curves (S. J. Perkins and R. B. Sim, unpublished results).

Acknowledgments

We thank Professor J. E. Fothergill for kindly providing the C1s sequence prior to publication and Dr. J. Ripoche and A. J. Day for the human factor H sequence.

Registry No. Factor H, 80295-65-4.

REFERENCES

Argos, P., & Mohanarao, J. K. (1986) *Methods Enzymol. 130*, 185–207.

Arlaud, G. J., & Gagnon, J. (1983) *Biochemistry 22*, 1758–1764.

Arlaud, G. J., Willis, A. C., & Gagnon, J. (1987) *Biochem. J. 241*, 711–720.

Ashida, T., Tanaka, I., & Yamane, T. (1981) *Int. J. Pept. Protein Res. 17*, 322–329.

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

Busetta, B., & Hospital, M. (1982) *Biochim. Biophys. Acta 701*, 111–118.

Byler, D. M., & Susi, H. (1986) *Biopolymers 25*, 469–487.

Caras, I. W., Davitz, M. A., Rhee, L., Weddell, G., Martin, D. W., & Nussenzweig, V. (1987) *Nature (London) 325*, 545–549.

Chothia, C. (1975) *Nature (London) 254*, 304–308.

Chothia, C., & Janin, J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4146–4150.

Chothia, C., & Janin, J. (1982) *Biochemistry 21*, 3955–3965.

Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry 13*, 211–222.

Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry 13*, 222–245.

Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135–175.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

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

Cid, H., Bunster, M., Arriagada, E., & Campos, M. (1982) *FEBS Lett. 150*, 247–254.

Cohen, F. E., Sternberg, M. J. E., & Taylor, W. R. (1981) *J. Mol. Biol. 148*, 253–272.

Cohen, F. E., Abarbanel, R. M., Kuntz, I. D., & Fletterick, R. J. (1986) *Biochemistry 25*, 266–275.

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

Day, A. J., Ripoche, J., Willis, A. C., & Sim, R. B. (1987) *Complement 4*, 147–148.

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

Eisenberg, D. (1984) *Annu. Rev. Biochem. 53*, 595–623.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Garratt, R. C., Taylor, W. R., & Thornton, J. M. (1985) *FEBS Lett. 188*, 59–62.

Haris, P. I., Lee, D. C., & Chapman, D. (1986) *Biochim. Biophys. Acta 874*, 255–265.

Hennessey, J. P., & Johnson, W. C. (1982) *Anal. Biochem. 125*, 177–188.

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

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

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

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

Kabsch, W., & Sander, C. (1983a) *FEBS Lett. 155*, 179–182.

Kabsch, W., & Sander, C. (1983b) *Biopolymers 22*, 2577–2637.

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.

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

Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H., & Fitch, W. M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3388–3392.

Lee, D. C., & Chapman, D. (1986) *Biosci. Rep. 6*, 235–256.

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

Leytus, S. P., Kurachi, K., Sakariassen, K. S., & Davie, E. W. (1986) *Biochemistry 25*, 4855–4863.

Lim, V. I. (1974a) *J. Mol. Biol, 88*, 857–872.

Lim, V. I. (1974b) *J. Mol. Biol. 88*, 873–894.

Lozier, J., Takahashi, N., & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3640–3644.

Mackinnon, C. M., Carter, P. E., Smyth, S. J., Dunbar, B., & Fothergill, J. E. (1987) *Eur. J. Biochem. 169*, 547–553.

Miller, J., Malek, T. R., Leonard, W. J., Greene, W. C., Shevach, E. M., & Germain, R. N. (1985) *J. Immunol. 134*, 4212–4217.

Milner-White, E. J., & Poet, R. (1986) *Biochem. J. 240*, 289–292.

Morley, B. J., & Campbell, R. D. (1984) *EMBO J. 31*, 153–157.

Nishikawa, K. (1983) *Biochim. Biophys. Acta 748*, 285–299.

Olinger, J. M., Hill, D. M., Jakobsen, R. J., & Brody, R. S. (1986) *Biochim. Biophys. Acta 869*, 89–98.

Perkins, S. J. (1986) *Eur. J. Biochem. 157*, 169–180.

Perkins, S. J., Chung, L. P., & Reid, K. B. M. (1986) *Biochem. J. 233*, 799–807.

Pouchert, C. J., Ed. (1981) *The Aldrich Library of Infrared Spectra*, Aldrich, Milwaukee, WI.

Reid, K. B. M. (1983) *Biochem. Soc. Trans. 11*, 1–12.

Reid, K. B. M. (1986) *Essays Biochem. 22*, 27–68.

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.

Ripoche, J., Day, A. J., Harris, T. J. R., & Sim, R. B. (1988) *Biochem. J. 249*, 593–602.

Rose, G. D. (1978) *Nature (London) 272*, 586–590.

Schulz, G. E., & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York.

Sim, E., Palmer, M., Puklavec, M., & Sim, R. B. (1983) *Biosci. Rep. 3*, 1119–1131.

Sim, R. B., & DiScipio, R. G. (1982) *Biochem. J. 205*, 285–293.

Susi, H. (1969) in *Structure and Stability of Biological Macromolecules* (Timasheff, S. N., & Fasman, G. D., Eds.) Vol. 2, pp 575–663, Dekker, New York.

Susi, H. (1972) *Methods Enzymol. 26*, 381–391.

Susi, H., & Byler, D. M. (1983) *Biochem. Biophys. Res. Commun. 115*, 391–397.

Susi, H., & Byler, D. M. (1986) *Methods Enzymol. 130*, 290–311.

Susi, H., Timasheff, S. N., & Stevens, L. (1967) *J. Biol. Chem. 242*, 5460–5466.

Tack, B. F., & Prahl, J. W. (1976) *Biochemistry 15*, 4513–4518.

Taylor, W. R., & Thornton, J. M. (1984) *J. Mol. Biol. 173*, 487–514.

Wasacz, F. M., Olinger, J. M., & Jakobsen, R. J. (1987) *Biochemistry 26*, 1464–1470.