Secondary Structure of a Complement Control Protein Module by Two-Dimensional ^{1}H NMR†

P. N. Barlow,‡ M. Baron,‡ D. G. Norman,‡ A. J. Day,§ A. C. Willis,§ R. B. Sim,§ and I. D. Campbell*,‡

Department of Biochemistry and MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Received August 27, 1990

ABSTRACT: The complement control protein (CCP) module (also known as the short consensus repeat) is a consensus sequence of about 60 amino acid residues which is thought to fold independently. It occurs over 140 times in more than 20 extracellular mosaic proteins including 12 proteins of the complement cascade. An isolated CCP module, the 16th repeat from human complement factor H, has been expressed in a yeast vector and shown to fold with the same pattern of disulfide bond formation as is seen in the native protein. Two-dimensional 600-MHz ^{1}H NMR spectra of this module have been recorded at pH 3.3 and 6.0 and analyzed to permit determination of secondary structure in solution. The CCP module comprises two predominantly extended segments (Glu1–His13 and Ala17–Glu27), two segments of double-stranded antiparallel β-sheet (Gly14–Val16 paired with Tyr31–Cys33 and Gly38–Asp40 paired with Ser57–Ile59), and a short piece of triple-stranded β-sheet (Glu27–Thr30, Ile44–Leu48, and Lys51–Ser53). Turns occur at Asp22, Gly36, and Glu50, while Gly41–Ala43 appear to form a looped-out segment or bulge. This structure is compared with a secondary structure prediction made on the basis of an alignment scheme of 101 sequences for CCP modules [Perkins, S. J., Haris, P. I., Sim, R. B., & Chapman, D. (1988) *Biochemistry 27*, 4004–4012]—the experimentally determined secondary structure bears an overall resemblance to the predicted one but differs in the number and position of turns. Some of those amino acid residues which are highly conserved throughout the range of CCP modules appear to play a role in stabilizing the global fold.

Mosaic proteins are built up from a number of different structural units or modules (Doolittle, 1985; Patthy, 1985; Baron et al., 1990a). Although they are considered to have evolved by mutation, duplication, insertion, and deletion of a common pool of such units, they exhibit a diverse range of biological functions. Mosaic proteins are involved in cell adhesion and migration (Williams & Barclay, 1988), embryogenesis (Wharton et al., 1985), and the pathways of blood clotting (Patthy, 1985; Furie & Furie, 1988), fibrinolysis (Patthy, 1985), and complement (Day et al., 1989; Sim & Perkins, 1989; Reid & Day, 1989). The modular units are defined by consensus sequences which often include conserved cysteines. In spite of the availability of sequence information, very little is known about the detailed tertiary structure of mosaic proteins. However, by determining the "consensus structure" of the limited number of module types, it will be possible to infer valuable information about a large number of proteins involved in a wide variety of biological systems. A strategy for achieving this goal, involving overexpression of an isolated module in a yeast vector followed by structure determination with NMR, has been successfully applied to the fibronectin type I module (Baron et al., 1990b). Below, we describe the application of the same techniques to another type of module, the complement control protein (CCP)[1] module (also known as the short consensus repeat).

The CCP module is about 60 amino acids in length, is generally encoded by a discrete exon at the gene level (Campbell et al., 1988), and has a characteristic consensus sequence involving four nearly invariant cysteines along with other highly conserved residues (Day, 1990). It is thought to fold independently with the framework cysteines forming intradomain disulfide bridges in the pattern Cys1–Cys3 and Cys2–Cys4 (Day et al., 1987; Janatova et al., 1989; Day, 1990). Between 2 and 30 CCP modules are present in each of 12 complement proteins (Reid & Day, 1989; Sim & Perkins, 1989). For example, complement factor H is an abundant plasma glycoprotein of M_r 155000 containing about 12% carbohydrate (Ripoche et al., 1988). Its amino acid sequence consists solely of 20 contiguous CCP modules. Complement receptors 1 and 2, C4b binding protein (C4bp), membrane cofactor protein, and decay accelerating factor are composed almost entirely of CCP modules arranged in contiguous fashion (Reid & Day, 1989). All of these proteins are encoded in the regulation of the complement activation gene cluster (Campbell et al., 1988) and share the property of interacting specifically with the complement proteins C3b and/or C4b. Furthermore, CCP modules are found in C2, factor B, C1r, C1s, C6, and C7—all proteins whose functional role includes binding to C3 or C3b, or to its homologues C4 and C5 (Day et al., 1989). It is tempting to ascribe a specific protein–protein recognition role to the CCP module; however while these proteins have many modules, only a few modules are implicated in binding C3b/C4b. The C3b binding site in factor H is likely to be localized within repeats 3–6 (Sim & Perkins, 1989) while in CR1 not more than 6 of the 30 CCP modules have been implicated in C3b or C4b binding (Klickstein et al.,

† This work is a contribution from the Oxford Centre for Molecular Science which is supported by the SERC and the MRC. P.N.B. was supported by a MRC postdoctoral training fellowship. A.J.D. thanks the trustees of the Schorstein Fellowship for their support.

‡ Department of Biochemistry.

§ MRC Immunochemistry Unit, Department of Biochemistry.

[1] Abbreviations: CCP, complement control protein; CCP16, 16th repeat (or CCP module) of factor H; DQF, double quantum filtered; NOE, nuclear Overhauser effect; HOHAHA, homonuclear Hartmann–Hahn.

0006-2960/91/0430-997$02.50/0

1988). Hence the CCP module may be thought of as a structural building block onto which function-related sequence variations are superimposed (Day et al., 1989).

Intriguingly, CCP modules are being identified in a growing number of noncomplement proteins (Day et al., 1989). These include blood clotting factor XIII b subunit, β2-glycoprotein I, interleukin 2 receptor, haptoglobin, the core protein from cartilage proteoglycan, and thyroid peroxidase. In addition, CCP modules have been identified in a recently described family of cell adhesion proteins, which are important in targeting white blood cells to their appropriate sites (Tedder et al., 1989; Bevilacqua et al., 1989; Springer, 1990). These include endothelial leukocyte adhesion molecule 1, lymphocyte-associated cell surface molecule 1, mouse lymph node homing receptor, and 140-kDa granule membrane protein. All these proteins consist of an N-terminal lectin domain, an epidermal growth factor-like module, a varying number of CCP modules, a transmembrane region, and a short cytoplasmic domain. In all, some 140 CCP modules have been found in over 20 proteins.

In the absence of X-ray crystallographic data, structural information for the CCP module has to date relied on neutron diffraction of C4bp (Perkins et al., 1986), electron microscopy of C4bp, CR1, and CR2 (Dahlback et al., 1983; Bartow et al., 1989; Moore et al., 1989), Fourier transform infrared spectroscopy of factor H, and a secondary structure prediction based on an alignment scheme for 101 CCP module sequences (Perkins et al., 1988). These studies were interpreted as being consistent with β-sheet-sandwich-type modules which are arranged in a "beads-on-a-string" fashion. In this paper we describe the use of protein engineering and solution-state high-resolution two-dimensional ^{1}H NMR spectroscopy to determine the detailed secondary structure of a representative CCP module, the 16th repeat from human factor H.

Materials and Methods

Strategy for Expression of the CCP Module. The 16th repeat from human complement factor H (CCP16) was expressed in a yeast secretion system. Yeast cells were transformed with the shuttle plasmid pMA91 containing coding sequence for the yeast α-factor leader peptide fused in frame with CCP16. The methodology for yeast expression has been described previously (Baron et al., 1990c). A phosphoglycerate kinase promoter (Mellor et al., 1983; Kingsman & Kingsman, 1985) and the α-factor leader (Kurjan & Herskowitz, 1982) were used to direct the expressed protein into the culture medium.

Construction of Vectors. The sequence of CCP16 [residues 909–967 in Ripoche et al. (1988)] is shown in Figure 2 (referred to here as residues 1–59) and is encoded by nucleotides 2852–3028 in the sequence of Ripoche et al. (1988). The amino acid numbering scheme used in Figure 2 will be employed hereafter. A 1171-bp *Bam*HI–*Hin*dIII fragment (nucleotides 2465–3635) from the human factor H specific cDNA clone R2a (Ripoche et al., 1988) was excised and subcloned into *Bam*HI/*Hin*dIII-cut M13mp18. The oligonucleotide 5′-AATCTGTTTATATGCAT (synthesized by Celltech Ltd., U.K.) was used with a mutagenesis kit (Amersham) to change nucleotide 3029 from an A to a T, introducing a stop codon after Ile59 of CCP16. A 215-bp *Stu*I–*Nco*I fragment (corresponding to nucleotides 2857–3072), containing the last nucleotide of Gly2 and the coding sequences for amino acids 3–59 of CCP16 and a stop codon, was then excised and cloned into pMB50 (Baron et al., 1990c) containing a unique polylinker with *Stu*I and *Nco*I restriction enzyme sites (underlined):

5′AGCTTGGACAAAAGAGAAGGCCTTCCTTCCATG

ACCTGTTTTCTCTTCCGGAAGGAAGGTACCTAG-5′

The vector pMB50 contains coding sequence for the yeast α-factor leader sequence with the last five amino acids of the α-factor peptide being encoded in frame with the first two amino acids of CCP16 in the polylinker. The *Stu*I restriction enzyme cuts between the second and third nucleotide of the Gly2 codon of CCP16 and thus allowed construction of coding sequence for the α-factor leader in frame with CCP16 followed by a stop codon.

The α-factor leader–CCP16 construct was then excised on a *Bgl*II–*Bam*HI fragment and cloned into the *Bgl*II site of the yeast expression vector pMA91 (Mellor et al., 1983) to give the expression plasmid pMA91/FHCCP16.

Yeast Transformation, Growth of Cells, and Purification. Competent yeast cells (*Saccharomyces cerevisiae* F107, leu2,3, pep4,3—a gift from Celltech Ltd.) were transformed with pMA91/FHCCP16 (Beggs, 1978; Hinnen et al., 1978). Transformants were selected by their ability to grow in the absence of leucine. One-liter batches of cells were routinely grown in 2.5-L flasks at 30 °C in an orbital shaking incubator. C18 reverse-phase beads (HPLC Technology Ltd., 2 g/L) were included in the media, which consisted of yeast nitrogen base without amino acids (Difco) supplemented at 0, 24, 36, and 48 h with 5 mL of 40% glucose and 5 mL of amino acid stock mix (without leucine). After 60 h the beads were separated from the cells and supernatants by filtration on a sintered glass funnel and then washed several times with water. Protein was removed from the beads by exhaustive washing with 50% acetonitrile. The protein solution was freeze-dried, redissolved in a small volume of 10% acetonitrile, and divided into two aliquots, and each aliquot was purified, with high-performance liquid chromatography, by two sequential passages through a reverse-phase column (Beckman C8, 250 × 100 mm) with a gradient running from 10 to 80% acetonitrile in 0.1% aqueous trifluoroacetic acid. Typically, 1–2 mg of purified protein was obtained from 20 L of cells. The purified CCP16 was subjected to N-terminal sequence analysis (see below) to check for correct expression and purity.

Disulfide Mapping and N-Terminal Analysis. Purified CCP16 (20 μg) was digested with endoproteinase Glu-C (Boehringer, sequencing grade) at a 1:20 enzyme:substrate weight ratio in 1% ammonium bicarbonate, pH 8.5, for 16 h at 37 °C. The resulting peptides were separated on a Vydac C4 column (250 × 4.6 mm) with a 0.1% trifluoroacetic acid/acetonitrile gradient, and a fraction of all peaks was analyzed for cysteic acid content after oxidation. Only the two major peaks had appreciable amounts of cysteic acid, and these were sequenced. Purified factor H CCP16 and disulfide-linked pairs of peptides generated by V8 digest of CCP16 were seqeunced by automated Edman degradation in an Applied Biosystems 470A protein gas-phase sequencer with on-line PTH analysis.

NMR Analysis. Purified CCP16 was dissolved in either D_2O or 90% H_2O/10% D_2O to give a concentration of 2–3 mM and the pH adjusted to uncorrected glass-electrode readings of either 3.3 or 6.0.

Solution-state ^{1}H NMR spectra were recorded on a Bruker AM 600 spectrometer at either 300 or 310 K. Double quantum filtered (DQF) correlation spectra (Piantini et al., 1982; Rance et al., 1983) and nuclear Overhauser effect (NOE) (Jeener et al., 1979; Kumar et al., 1981) spectra were recorded in a phase-sensitive manner by the time proportional phase increment method and CYCLOPS with a mixing time

of 250 ms. Homonuclear Hartmann–Hahn (HOHAHA) (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) spectra were collected in reverse mode with transfer of net magnetization by the WALTZ17$_y$ mixing sequence (Bax et al., 1987). The mixing time was 50 ms. In the case of spectra recorded in D_2O, irradiation of the solvent resonance was carried out during the relaxation delay and, for the NOE spectra, the mixing time. Lower decoupler powers, just sufficient to maintain saturation, were applied during the t_1 periods. In the case of spectra recorded in 90% H_2O/10% D_2O, a "jump–return" sequence (Plateau & Gueron, 1982) was used to suppress the solvent resonance. The receiver phase was adjusted to eliminate base-line distortions (Marion & Bax, 1988). Prior to Fourier transform in F_2 the free induction decays were deconvoluted in the time domain with a Gaussian function (Driscoll et al., 1989). This eliminates t_2 ridges in the final spectrum.

Amide hydrogens which were relatively slow to exchange with solvent were identified by incubating protonated CCP16 in D_2O for 18 h at room temperature and then recording a DQF correlation spectrum over 6 h.

Results

Expression and Characterization of Factor H CCP16. Amino-terminal sequence analysis confirmed that the expressed CCP16 had a sequence identical with that of repeat 16 of native factor H (Ripoche et al., 1988) with the N-terminal 56 out of the 59 amino acid residues being positively identified. To assess whether the expressed module had the correct formation of disulfide bonds, purified CCP16 was digested with endoproteinase Glu-C and the disulfide-linked peptide pairs were purified (results not shown) and subjected to N-terminal sequence analysis. Peptide EGLP-KSPPEISHGVVAHMSDSYQYGEE (where the dash refers to Cys5) was linked to peptide GFGIDGPAIAK-L(G)E (Cys47), and peptide VTYK-FE (Cys33) was linked to peptide KWSHPPS-I (Cys58). This is consistent with the disulfide bond organization in repeat 16 of native factor H where it has been shown that Cys5 is disulfide bonded to Cys47 (Day, 1990). This arrangement of disulfide bridges is believed to be a general feature of the CCP module (see introduction); i.e., the first Cys is disulfide bonded to the third Cys and the second Cys is disulfide bonded to the fourth. The disulfide mapping procedure provided further confirmation that expressed CCP16 has a sequence identical with that of repeat 16 of native factor H.

NMR Analysis of CCP16. Spectra collected at 310 K and pH 3.3 were assigned in detail, although data collected at 300 K, pH 3.3 (data not shown), and at 310 K, pH 6.0, were of assistance in cases of overlapping resonances. Sequential assignment was carried out according to the methods of Wüthrich et al. (1982); spin systems were defined from NOE, HOHAHA, and DQF-correlated spectroscopy and connected by means of sequential NOEs, and the resulting sequence of spin systems was equated with a unique portion of the primary sequence. The sequence of CCP16 includes six proline residues for which β_i–δ_{i+1} connectivities were used to establish sequential assignments. Figure 1 illustrates the connection of spin systems for two segments of the protein.

By this means it was possible to assign nearly all the spin systems to a unique position in the sequence and hence to compile a virtually complete list of assignments at pH 3.3 (Table I). Difficulty was encountered in tracing the side chain of Lys32 due to overlap of the resonance from its α-protons with those of several other residues and the difficulty of detecting amide–side chain cross-peaks in the NOE spectra.

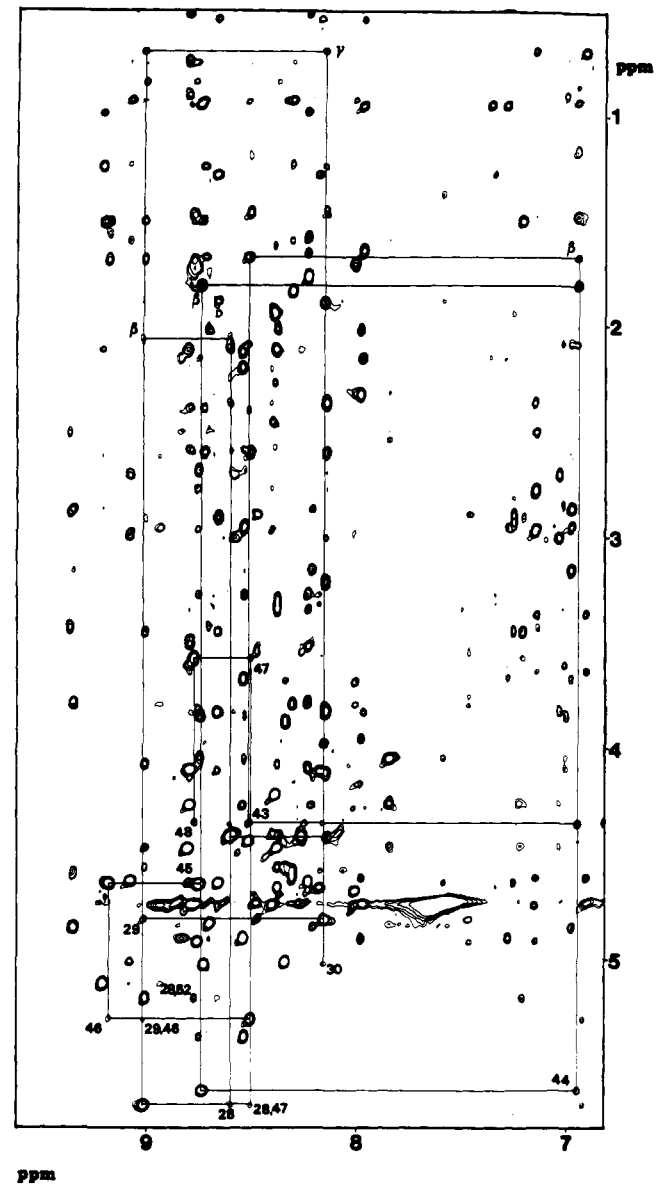

FIGURE 1: Part of a NOESY spectrum of CCP16 in 90% H_2O/10% D_2O at pH 3.0 and 310 K with a mixing time of 250 ms. The region shown is that containing some of the cross-peaks between the backbone NH (F_2 axis) and α, β, and γCH (F_1 axis) resonances. Two sets of sequential NOE connectivities are indicated by unbroken lines with the labels at the intraresidue αCH–NH cross-peaks. Notice that βCH_i–NH_{i+1} and γCH_i–NH_{i+1} cross-peaks (labeled β and γ, respectively) were employed along with the αCH_i–NH_{i+1} cross-peaks to obtain sequential assignments. Some long-range αCH_i–NH_j cross-peaks are also labeled.

Identification of its δ- and ϵ-protons was made by a process of elimination. The ϵ-protons of Lys49 were not detected, probably due to exchange broadening. It was not possible to identify the C4 protons of Phe34, and it is likely that this resonance is coincident with that due to either the C2,6 or the C3,5 protons of Phe34. Assignment of resonances from Glu1 was accomplished by a process of elimination since no connectivities from the N-terminal residue to the rest of the molecule were identified. A partial assignment was carried out at pH 6.0 (Table I). This not only assisted the assignment at pH 3.3 but also provided evidence for the conformational stability of the module over a 3-unit pH range.

The sequential NOEs observed in the spectrum of CCP16 are summarized in Figure 2 together with those amide protons that are slowly exchanging and values of the coupling constant $^3J_{HN\alpha}$ that could reliably be measured. Two of the gaps (at

Table I: Assignments of Resonances in CCP16 at 310 K and pH 3.3 and 6.0[a]

	pH 3.3						pH 6.0					
residue	NH	αH	βH	γH	others		NH	αH	βH	γH	others	
Glu1		4.06	2.24	2.62			–	–	–	–		
			2.24	2.62					–	–		
Gly2	8.23	3.26					8.20	3.20				
		3.78						3.68				
Leu3	8.23	4.65	1.57	1.77	δH	0.91	8.20	4.66	1.55	1.74	δH	0.95
			1.65			0.98			1.64			0.95
Pro4		4.90	2.14	1.92	δH	3.69		–	–	1.92	δH	3.69
			2.40	2.11		3.87			–	2.11		3.87
Cys5	8.83	4.83	2.49				–	4.83	2.49			
			3.43						3.43			
Lys6	8.70	4.90	1.99	1.66	δH	1.90	8.75	4.91	1.99	–	δH	–
			2.33	1.73		1.90			2.33	–		–
					εH	3.14					εH	–
						3.14						–
					εNH	7.55					εNH	–
Ser7	7.98	4.61	3.95				8.01	4.56	3.94			
			4.17						4.15			
Pro8		2.87	−0.11	1.11	δH	3.35		2.87	−0.13	1.07	δH	3.36
			1.25	1.46		3.63			1.25	1.42		3.62
Pro9		4.20	1.61	1.79	δH	2.80		4.20	1.58	1.79	δH	2.80
			2.27	1.84		2.96			2.23	1.84		2.96
Glu10	8.37	4.47	1.93	2.46			8.37	4.43	1.89	2.42		
			2.11	2.56					2.13	2.42		
Ile11	8.80	4.63	1.77	0.75	γCH_3	0.53	8.88	4.60	–	0.72	γCH_3	0.50
				0.90	δCH_3	0.31				0.87	δCH_3	0.27
Ser12	7.80	4.08	3.34				–	–	–			
			3.52									
His13	8.79	3.61	2.07		H2	8.62	–	–	–		H2	
			2.86		H4	7.01					H4	7.00
Gly14	6.99	3.82					6.96	3.83				
		4.07						3.96				
Val15	8.74	4.73	2.15	0.95			8.71	–	2.13	0.95		
				0.95						0.95		
Val16	7.96	3.83	1.63	0.19			7.99	3.81	1.59	0.19		
				0.55						0.50		
Ala17	8.65	4.10	1.29				8.59	4.05	1.27			
His18	8.16	4.65	3.26		H2	8.69	8.06	–	3.23		H2	–
			3.30		H4	7.33			3.30		H4	7.23
Met19	8.35	4.40	2.01	2.64	εH	2.16	8.42	4.38	1.98	2.61	εH	2.14
			2.15	2.64					2.09	2.61		
Ser20	8.25	4.72	3.48				8.10	–	3.39			
			3.35						3.39			
Asp21	8.44	4.80	2.88				8.42	–	2.77			
			2.88						2.77			
Ser22	7.46	4.91	3.73				7.38	4.92	3.72			
			3.77						3.72			
Tyr23	8.75	4.76	2.76		H2,6	7.14	8.76	–	2.78		H2,6	7.17
			2.95		H3,5	6.90			2.94		H3,5	6.86
Gln24	8.40	4.40	1.93	2.20	εH	6.86	9.06	4.37	1.96	–	εH	–
			2.37	2.20		7.57			2.26	–		–
Tyr25	8.60	3.79	2.69		H2,6	7.01	8.69	3.72	2.63		H2,6	7.00
			3.00		H3,5	6.80			2.95		H3,5	6.79
Gly26	8.15	3.20					8.16	3.12				
		4.10						4.07				
Glu27	8.15	4.41	2.36	2.67			8.17	4.28	2.36	2.51		
			2.36	2.77					2.36	2.62		
Glu28	8.60	5.69	2.04	2.25			8.49	5.58	1.98	2.50		
			2.10	2.60					2.10	2.50		
Val29	9.01	4.83	1.89	0.45			9.04	4.85	–	0.50		
				0.69						0.66		
Thr30	8.21	5.02	3.97	1.25			8.09	4.99	3.97	1.22		
Tyr31	8.72	4.74	2.38		H2,6	6.49	8.66	–	2.33		H2,6	6.49
			2.61		H3,5	6.66			2.56		H3,5	6.64
Lys32	8.74	4.64	–	–	δH	1.75	8.68	4.60	1.71	1.40	δH	1.71
						1.75			1.79	1.40		1.71
					εH	3.07					εH	3.00
						3.07						3.00
					εNH	7.54					εNH	–
Cys33	8.74	5.36	2.67				8.73	5.34	2.67			
			3.26						3.24			
Phe34	8.55	4.90	2.95		H2,6	7.27	8.54	4.88	2.92		H2,6	7.27
			3.66		H3,5	7.34			3.65		H3,5	7.34
					H4	–					H4	–
Glu35	8.55	4.25	2.31	2.59			8.57	4.21	2.07	2.41		
			2.39	2.71					2.13	2.41		

Table I (Continued)

	pH 3.3						pH 6.0					
residue	NH	αH	βH	γH	others		NH	αH	βH	γH	others	
Gly36	8.73	3.50					8.77	3.48				
		4.15						4.08				
Phe37	8.20	4.85	2.85		H2,6	6.97	8.18	4.82	3.10		H2,6	6.97
			3.15		H3,5	6.86			3.10		H3,5	–
					H4	7.04					H4	–
Gly38	9.34	3.80					9.67	3.75				
		4.58						4.59				
Ile39	8.23	4.62	1.82	1.23	γCH_3	0.93	8.27	4.63	1.82	1.23	γCH_3	0.91
				1.59	δCH_3	0.74				1.59	δCH_3	0.73
Asp40	9.06	5.01	2.69				9.00	4.90	2.60			
			2.99						2.76			
Gly41	8.34	3.86					8.13	3.66				
		4.66						4.51				
Pro42		4.42	2.08	2.23	δH	3.67		4.40	2.05	–	δH	3.67
			2.39	2.23		3.90			2.22	–		–
Ala43	8.51	4.35	1.67				8.51	4.33	1.65			
Ile44	6.92	5.62	1.80	1.17	γCH_3	0.93	6.91	5.58	1.78	1.15	γCH_3	0.92
				1.47	δCH_3	0.84				1.43	δCH_3	0.84
Ala45	8.73	4.63	1.50				8.70	4.59	1.46			
Lys46	9.17	5.28	1.54	1.15	δH	1.63	9.18	5.29	1.48	1.07	δH	–
			1.67	1.19		1.63			1.66	1.16		–
					εH	2.91					εH	–
						2.91						–
					εNH	–					εNH	–
Cys47	8.50	3.57	1.47				8.59	3.48	1.48			
			2.60						2.62			
Leu48	8.77	4.34	1.44	1.69	δCH_3	0.74	8.76	4.31	1.47	1.70	δCH_3	0.71
			1.44			0.84			1.47			0.80
Gly49	7.19	4.26					7.13	4.34				
		4.46						4.48				
Glu50	7.84	4.04	2.31	2.53			7.66	3.88	2.32	–		
			2.47	2.53					2.40	–		
Lys51	7.58	4.62	1.88	1.27	δH	1.71	7.75	4.60	1.84	1.27	δH	–
			1.94	1.37		1.71			1.94	1.37		–
					εH	3.01					εH	–
						3.01						–
					εNH	7.60					εNH	–
Trp52	8.65	5.17	2.89		2H	7.23	8.59	5.14	2.87		2H	7.22
			3.44		4H	7.19			3.41		4H	7.18
					5H	6.90					5H	6.90
					6H	6.11					6H	6.03
					7H	7.14					7H	7.07
					NH	9.35					NH	9.35
Ser53	9.00	4.45	4.06				9.01	4.44	4.03			
			4.54						4.53			
His54	8.37	4.64	3.26		H2	8.69		–	3.28		H2	–
			3.31		H4	7.48			3.28		H4	7.46
Pro55		4.38	1.54	1.65	δH	3.33		4.38	1.53	1.11	δH	3.22
			1.72	1.72		3.87			1.71	2.47		3.22
Pro56		4.67	1.75	1.12	δH	2.48		–	1.63	–	δH	3.30
			2.35	1.12		3.24			1.70			3.85
Ser57	7.99	4.74	3.66				8.05	–	3.63			
			3.78						3.76			
Cys58	8.93	5.11	2.61				8.82	5.08	2.57			
			2.95						2.90			
Ile59	9.20	4.70	2.11	1.22	γCH_3	0.98	9.26	4.61	2.06	1.19	γCH_3	0.94
				1.22	δCH_3	1.50				1.19	δCH_3	1.47

[a] Chemical shifts are with respect to dioxane; (–) indicates that the resonance was not observed or that its position was not established.

Gly2 and Gly26) arise where consecutive amide protons had coincident resonances under all conditions. The other gap arises because no connectivity could be detected between His54 and Pro55. Strong α_i–N_{i+1} NOEs and large (>8 Hz) values of $^3J_{HN\alpha}$ are considered evidence for extended segments of peptide, turns are identified from strong N_i–N_{i+1} NOEs, and regions of α-helix give rise to stretches of consecutive, strong NH_i–NH_{i+1} NOEs and small $^3J_{HN\alpha}$ coupling constants. A qualitative inspection of the data summarized in Figure 2 indicates that CCP16 is predominantly extended with discrete turns at Asp21, Gly36, and Glu50. Taken together with the run of four slowly exchanging amides from Ile44–Cys47 and those at Thr30, Lys32, and Ile59, this is evidence for a predominance of turns and β-sheet in the region Val15–Ile59. The segment bounded by Glu1–Gly14 does not seem to conform to a classical structural motif; however, the slowly exchanging amide at Ile11 is consistent with a defined fold while the very upfield shifted βH of Pro8 implies close proximity to an aromatic side chain and hence an intimate association between this region and the β-sheet-rich region. There is no evidence for α-helical structure anywhere in the molecule.

In order to define further the secondary structure of the CCP module, long-distance backbone–backbone NOEs were considered in conjunction with the sequential ones. As illustrated in Figure 3, these NOEs are consistent with two segments of double-stranded antiparallel β-sheet (Gly14–Val16

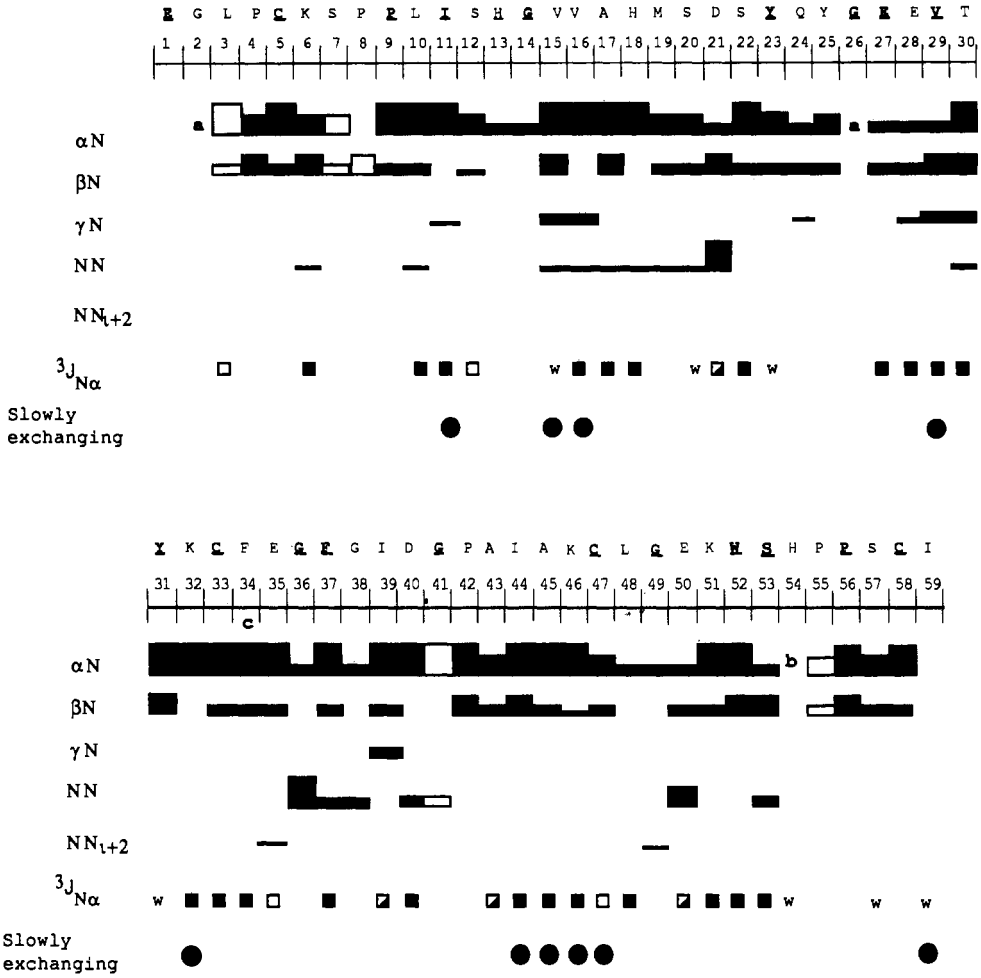

FIGURE 2: Summary of highly conserved residues in the CCP module, the sequential NOEs observed for CCP16, the $^3J_{HN\alpha}$ coupling constant data, and the slowly exchanging amides. The bar heights correspond approximately to the NOE intensities. White bars are used when the NOE is to the δH of proline. Standard notation is used for the NOE classes. Black circles indicate those amide protons considered to be slowly exchanging, i.e., those that are detected in a DQF correlation spectrum 16–24 h after dissolution in D_2O. Those residues labeled in boldface and underlined are considered to be highly conserved (Ripoche et al., 1988) within CCP modules. (a) The amide proton resonance could not be resolved from that of the subsequent residue. (b) No connectivity to the subsequent proline was identified. (c) The amide protons of F34 and E35 were coincident at pH 3.3 but resolved at pH 6.0; hence, the NOEs displayed in this case are at pH 6.0—all others are at pH 3.3. The symbols used for the coupling constant data are as follows: white squares, $^3J_{HN\alpha} < 7$ Hz; half black squares, $^3J_{HN\alpha} = 7$–8 Hz; black squares, $^3J_{HN\alpha} > 8$ Hz; w, overlap with residual water peak.

paired with Tyr31–Cys33 and Gly38–Asp40 paired with Ser57–Ile59) and a short piece of triple-stranded β-sheet (Glu27–Thr30, Ile44–Leu48, and Lys51–Ser53). As indicated in Figure 3, the two strands connecting the triple-stranded sheet with the double-stranded sheet at the C-terminus (i.e., Gly41–Ala45 and His54–Pro56) differ in length, and it seems likely that residues 41–43 are looped out, perhaps constituting a β-bulge-like structure. The region between Ala17 and Glu28 presumably forms an extended loop.

Discussion

The secondary structure of a CCP module has been elucidated with two-dimensional ^1H NMR. The 16th repeat of human factor H was chosen for this study on the basis of the following criteria: (a) its amino acid sequence is representative of the CCP modules in general; (b) the position of the consensus cysteine 1 to cysteine 3 disulfide linkage (residues 5 and 47) has been established in the native protein (Day, 1990); (c) it contains no Asn-linked glycosylation site; (d) the flanking cDNA sequence has appropriate restriction sites for vector construction.

The 59 amino acid protein, expressed and secreted from a yeast vector, folds to give a single conformation which has the correct formation of disulfide bonds. This observation implies that the structure of the isolated CCP module will be very similar to or identical with that of the module in the native protein. It thus provides strong support for the contention that CCP modules fold independently within the native protein and is consistent with calorimetric measurements made on unfolding of the CCP modules in C1s (Medved et al., 1989).

The first 13 residues of the CCP module form an extended strand which has a defined fold. The remainder of the molecule is rich in β-sheet and turns with 23 out of the 46 residues involved in sheet formation and at least three turns. Residues Ala17–Gly26 constitute a poorly defined loop, and Gly41–Ala43 appear to lie in a small looped-out segment or bulge. This secondary structure information, in conjunction with the pattern of disulfide bond formation, implies that the CCP module is compact and globular. The extended segment, Glu1–His13, is anchored at one end to the triple-stranded β-sheet by the disulfide 5–47 and at the other end to the β-sheet formed from Gly14–Val16 and Tyr31–Cys33. This segment of β-sheet is, in turn, disulfide bonded to the C-terminal β-sheet comprising Gly38–Asp40 and Ser57–Ile59. These secondary structural elements are conserved over the pH range 3.3–6.0 at 310 K. It is known that factor H and other complement proteins that contain CCPs are fully active at pH 6.0.

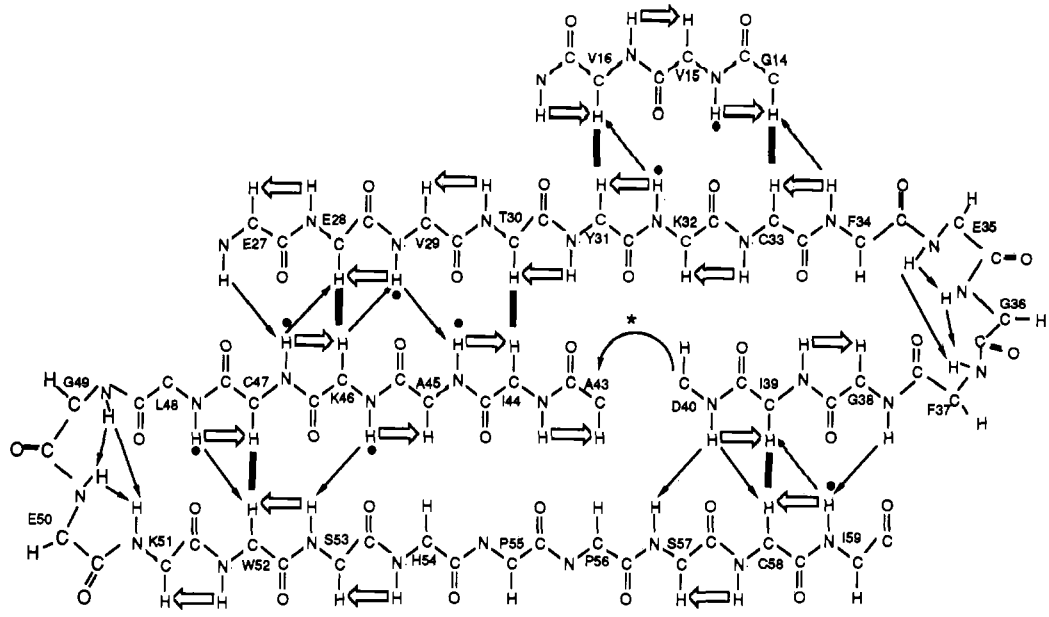

FIGURE 3: Secondary structural elements of CCP16 deduced from a qualitative interpretation of the NOE and slowly exchanging amide data. The peptide backbone is labeled at the α-carbons with the residue number and single-letter code. Thick black lines represent cross-sheet $CH\alpha_i$–$CH\alpha_j$ NOEs. Thin straight black arrows represent NH_i–NH_j and NH_i–$CH\alpha_j$ NOEs. The thick white arrows represent NH_i–$CH\alpha_{i+1}$ NOEs. Black circles indicate those amide protons which are slowly exchanging. (*) This curved arrow represents Gly41 and Pro42 which have been omitted from the diagram for the purpose of clarity.

Since CCP16 is a representative CCP module, it is interesting to compare its experimentally determined secondary structure with the generalized secondary structure predicted (Perkins et al., 1988) on the basis of an alignment scheme for 101 sequences for the CCP module as found in factor H and 12 other proteins. Perkins et al. predicted a high incidence of β-sheet in the region bounded by residues 23 and 59 (numbered on the basis of alignment with CCP16), with strands of approximately 7 residues interpolated by turns. Predictions for residues 4–22 were more tentative but gave some support for an extended segment containing three turns. These predictions were consistent with Fourier transform infrared spectra of factor H, which indicated a high content of β-sheet. In general, the gross features of the predicted structure are in reasonable agreement with the experimental data for CCP16 presented here. There is less agreement on the number and positions of turns and hence the length of strands and loops. Predictions of turns at positions 21, 36, and 50 are borne out by the experimental data. However, the predicted turn at residue 42 is probably a looped-out segment or bulge, while that at residue 55 appears to be inconsistent with the observed long-distance backbone–backbone NOEs (see Figure 3). The NOE data do not address the turn predicted at 27 because of the coincidence of resonances for the amide protons of Gly26 and Glu27. The structure of CCP16 thus appears to be less regular than was predicted. However, its globular nature is consistent with the "beads-on-a-string" model for CR1 derived from electron microscopy (Klickstein et al., 1988; Bartow et al., 1989).

The CCP module is characterized by a number of highly conserved residues (Figure 2). Many would appear to play an important structural role—obviously this is true for the cysteines and appears also to be the case for the glycines which occur mainly in turns. The conserved hydrophobic residues Trp52, Tyr31, and Val29 all occur within the β-sheet-rich region and may well be involved in forming a hydrophobic core. The determination of the refined tertiary structure, based on a full set of long-distance NOEs and molecular modeling, is ongoing in this laboratory and will enable a much fuller description of the spatial arrangement of the conserved residues.

As was pointed out in the introduction, it is improbable that all of the 20 CCP modules in factor H are involved in the specific interaction with C3b which characterizes many of the CCP-containing proteins. Factor H has other activities which involve binding to specific receptors on leukocytes and interaction with multiple charged groups on complement activating surfaces (Erdei & Sim, 1987; Pangburn & Meri, 1989). Therefore, it is possible that several of the modules are involved in these interactions. The remaining modules may play a structural or communicative role. In this context, it will be interesting to study the extent and nature of the interface between two CCP modules since this will define the rigidity of the protein. Efforts are currently underway toward the structure determination of an isolated double module in this laboratory.

It is likely that conclusions derived from factor H will be applicable to the wide range of other CCP module containing proteins. Hence, the application of two-dimensional NMR techniques to isolated modules is proving a successful strategy for gaining structural insight into previously inaccessible problems.

ACKNOWLEDGMENTS

We thank Paul Driscoll, Denise Mulvey, and Kenneth Reid for their advice and encouragement.

Registry No. Complement factor H, 80295-65-4.

REFERENCES

Baron, M., Norman, D. G., & Campbell, I. D. (1990a) *Trends Biochem. Sci.* (in press).

Baron, M., Norman, D., Willis, A., & Campbell, I. D. (1990b) *Nature 345*, 642–646.

Baron, M., Kingsman, A., Kingsman, S. M., & Campbell, I. D. (1990c) in *Protein Production in Biotechnology* (Harris, T. J. R., Ed.) pp 49–60, Elsevier, London.

Bartow, T., Klickstein, L. B., Wong, W., Roux, K., & Fearon, D. T. (1989) *FASEB. J. 3*, A501.

Bax, A., Sklenar, V., Clore, G. M., & Gronenburg, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511–6513.

Beggs, J. D. (1978) *Nature 275*, 104-109.
Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A., & Seed, B. (1989) *Science 243*, 1160-1165.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521-528.
Campbell, R. D., Law, S. K. A., Reid, K. B. M., & Sim, R. B. (1988) *Annu. Rev. Immunol. 6*, 161-195.
Dahlback, B., Smith, C. A., & Muller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3461-3465.
Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820-2821.
Day, A. J. (1990) in *Biochemistry and Molecular Biology of Complement* (Sim, R. B., Ed.) Kluwer Academic Publishers, Lancaster, U.K. (in press).
Day, A. J., Ripoche, J., Willis, A. C., & Sim, R. B. (1987) *Complement 4*, 147-148.
Day, A. J., Campbell, R. D., & Reid, K. B. M. (1989) in *Progress in Immunology* (Melchers, F., et al., Eds.) Vol. VII, pp 209-212, Springer-Verlag, Heidelberg.
Doolittle, R. F. (1985) *Trends Biochem. Sci. 114*, 233-237.
Driscoll, P. C., Clore, G. M., Beress, L., & Gronenburg, A. M. (1989) *Biochemistry 28*, 2178-2187.
Erdei, A., & Sim, R. B. (1987) *Biochem. J. 246*, 149-156.
Furie, B., & Furie, B. C. (1988) *Cell 53*, 505-518.
Hinnen, A., Hicks, J. B., & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1929-1933.
Janatova, J., Reid, K. B. M., & Willis, A. C. (1989) *Biochemistry 28*, 4754-4761.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Phys. Chem. 71*, 4546-4553.
Kingsman, A. J., & Kingsman, S. M. (1985) *Biotech. Genet. Eng. Rev. 3*, 377-416.
Klickstein, L. B., Bartow, T., Miletic, V., Rabson, L. D., Smith, J. A., & Fearon, D. T. (1988) *J. Exp. Med. 168*, 1699-1717.
Kumar, A., Ernst, R. R., & Wuthrich, K. (1981) *Biochem. Biophys. Res. Commun. 95*, 1-6.
Kurjan, J., & Herskowitz, I. (1982) *Cell 30*, 933-943.
Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528-533.
Medved, L. V., Busby, T. F., & Ingham, K. C. (1989) *Biochemistry 28*, 5408-5414.
Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J., & Kingsman, S. M. (1983) *Gene 24*, 1-14.
Moore, M. D., DiScipio, R. G., Cooper, N. R., & Nemerow, G. R. (1989) *J. Biol. Chem. 264*, 20576-20582.
Pangburn, M. K., & Meri, S. (1989) *Complement Inflamm. 6*, 383.
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.
Piantini, U., Sorensen, O. W., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 104*, 6800-6801.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310-7311.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479-485.
Reid, K. B. M., & Day, A. J. (1989) *Immunol. Today 10*, 177-180.
Ripoche, J., Day, A. J., Harris, T. J. R., & Sim, R. B. (1988) *Biochem. J. 249*, 593-602.
Sim, R. B., & Perkins, S. J. (1989) *Curr. Top. Microbiol. Immunol. 153*, 99-122.
Springer, T. A. (1990) *Nature 346*, 425-434.
Tedder, T. F., Isaacs, C. M., Ernst, T. J., Demetri, G. D., Adler, D. A., & Disteche, C. M. (1989) *J. Exp. Med. 170*, 123-133.
Wharton, K. A., Johansen, K. M., Xu, T., & Artavansis-Tsakonas, S. (1985) *Cell 43*, 567-581.
Williams, A. F., & Barclay, A. N. (1988) *Annu. Rev. Immunol. 6*, 381-405.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311-319.

Mössbauer Spectroscopy of Iron–Ovotransferrin: A Crystal Field Interpretation†

Kevork Spartalian,*,‡ George Lang,§ and Robert C. Woodworth‖
Departments of Biochemistry and Physics, University of Vermont, Burlington, Vermont 05405, and Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802

Received July 23, 1990; Revised Manuscript Received October 9, 1990

ABSTRACT: Mössbauer spectra from frozen solutions of ovotransferrin were recorded in a variety of applied external magnetic fields and at various temperatures in a small applied field. The results were fitted to a simple model for the electronic structure at the iron site. This model requires admixtures of the free ion ^6S and ^4P states, indicating a weak cubic crystal field. Possible implications of this model regarding the binding site are discussed.

The transferrins, a series of ca. 80-kDa glycoproteins found in the blood plasma and secretions of all vertebrates, serve a primary physiological role of sequestering and detoxifying ferric iron and, in the case of the plasma type, transporting iron to those tissues and cells of the body requiring iron (Aisen & Listowsky, 1980; Chasteen, 1983; Ponka et al., 1990). The X-ray crystal structures of two of these proteins reveal that each comprises two conformationally similar lobes connected by a short peptide sequence, each lobe being made up of two domains which define a deep cleft (Anderson et al., 1987, 1989; Baker et al., 1987; Bailey et al., 1988). Each cleft contains the amino acid side chains required to sequester one ferric ion and one synergistic anion, normally carbonate, essential for

† Supported by USPHS Grant DK21739 and an NSF EPSCoR Grant.
* Author to whom correspondence should be addressed.
‡ Department of Physics, University of Vermont.
§ Department of Physics, The Pennsylvania State University.
‖ Department of Biochemistry, University of Vermont.

0006-2960/91/0430-1004$02.50/0