Sheetz, M. P., and Chan, S. I. (1972), *Biochemistry 11*, 4573.

Singleton, W. S., Gray, M. S., Brown, M. L., and White, J. L. (1965), *J. Amer. Oil. Chem. Soc. 42*, 53.

Träuble, H., and Sackmann, E. (1972), *J. Amer. Chem. Soc. 94*, 4499.

Wang, J. H., and Copeland, E. (1973), *Proc. Nat. Acad. Sci. U. S. 70*, 1909.

Variable Region Sequence of the Heavy Chain from a Phosphorylcholine Binding Myeloma Protein†

Stuart Rudikoff* and Michael Potter

ABSTRACT: The variable region sequence of the heavy chain from McPC 603, a phosphorylcholine binding myeloma protein, has been determined primarily by use of the automated sequencer. The variable region of this protein contains methionine residues at positions 34 and 83. Three cyanogen bromide fragments were isolated from cleaved heavy chains and pepsin Fab's which accounted for this entire sequence. The sequence of this protein outside the hypervariable regions shows considerable homology to the variable regions of other mouse as well as human proteins suggesting a conservation of genes coding for heavy chains.

Mouse myeloma proteins with antigen binding specificity for a wide variety of antigens such as phosphorylcholine (Potter and Leon, 1968; Potter and Lieberman, 1970; Sher *et al.*, 1971), β-(1→6)-D-galactan (Potter *et al.*, 1972; Jolley *et al.*, 1973; Rudikoff *et al.*, 1973), α-(1→3)-dextran (Leon *et al.*, 1970; Weigert *et al.*, 1970), α-(1→6)-dextran, β-(2→1)-fructosan (Cisar *et al.*, 1974), and dinitrophenol (Eisen *et al.*, 1968; Jaffe *et al.*, 1969) have been previously described. These proteins provide excellent models for studying antibody structure as well as for exploring structure-function relationships among proteins binding the same hapten.

We have begun to explore in depth the structural and functional properties of a group of five phosphorylcholine binding proteins (M603, M167, T15, S107, and H8) all of which originated in the highly inbred BALB/c strain of mice. Many of the questions relating to the structure of binding sites depend upon a determination of both three-dimensional and primary structures. A systematic attempt has been made to crystallize the pepsin Fab fragments from our collection of phosphorylcholine binding proteins. Crystals suitable for X-ray diffraction studies were obtained from the Fab fragment of M603 and have been previously described (Rudikoff *et al.*, 1972). Padlan *et al.* (1973) have extended these initial observations and recently described the three-dimensional structure of this molecule at 4.5-Å resolution.

In the present study we have determined the variable region sequence from the heavy chain of M603. We propose to use this sequence in the construction of a three-dimensional model of the M603 Fab and as a prototype in comparative studies with the other phosphorylcholine binding proteins described above.

Materials and Methods

Protein Purification. Plasmacytoma McPC 603 (IgA, κ) has previously been described by Potter and Leon (1968) and Leon and Young (1971). The protein (M603) was purified by affinity chromatography on Sepharose-phosphorylcholinecolumns as described by Chesebro and Metzger (1972).

Heavy Chain Preparation. M603 protein (20–30 mg/ml) was dialyzed against 0.15 M Tris-HCl–0.15 M NaCl–2 mM Na_2EDTA and was reduced with 10 mM dithiothreitol for 2 hr at room temperature, followed by alkylation for 15 min with 20 mM iodoacetamide (Bridges and Little, 1971). The partially reduced and alkylated protein was dialyzed overnight against 6 M urea–1 M acetic acid, and heavy and light chains were separated by chromatography on a Sephadex G-100 column (5 × 100 cm) equilibrated in 6 M urea–1 M acetic acid.

Pepsin Fragments. Pepsin Fab's were prepared as previously described (Rudikoff *et al.*, 1972). Protein, partially reduced and alkylated as described above, was dialyzed against 0.1 M sodium acetate (pH 4.5) and digested with pepsin (Worthington) at a weight ratio of 1:100 (enzyme:protein) for 6 hr at 37°. The digestion was stopped by adjusting the pH to 8.6 by the addition of 2 M Tris and the Fab was separated by chromatography on Sephadex G-100 columns equilibrated in borate-buffered saline (pH 8.0).

Cyanogen Bromide Cleavage. Proteins were dissolved in 70% formic acid and CNBr was added at a 4:1 weight ratio (CNBr:protein). The reaction mixture was allowed to stand overnight at 4° and was then diluted with water and lyophilized. Fragments derived from cleavage of heavy chains will be denoted as Cn while those derived from cleavage of the pepsin Fab will be designated Cn′.

Sequence Determination. Amino acid compositions were determined on a Beckman 119 amino acid analyzer equipped with high sensitivity cuvets and recorder following hydrolysis of peptides in 6 N HCl for ~18 hr in evacuated and sealed tubes. Automated sequence determinations were performed on a Beckman Model 890C sequencer using the standard dimethylallylamine peptide program. Sequencer fractions obtained after each degradation cycle were converted to phenylthiohydantoin derivatives (Pth)[1] as previously described (Rudi-

†From the Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014. *Received April 11, 1974.*

[1]Abbreviation used is: Pth, phenylthiohydantoin.

koff *et al.*, 1973). Pth derivatives were identified by gas chromatography (gc) on SP400 columns as described by Pisano *et al.* (1972) and/or by amino acid analysis following hydrolysis of the Pth derivative to the free amino acid according to the hydriodic acid method of Smithies *et al.* (1971). In the above systems the Pth derivatives of Ala, Ser, Gly, Val, Pro, Thr, Leu-Ile, Met, Phe, Tyr, and Trp are identified by gc. Positions not identified by gc were hydrolyzed along with the preceding step to serve as a control for background and subjected to amino acid analysis. All of the Pth derivatives are identifiable as their respective free amino acids following hydrolysis with the exceptions of Met which is destroyed, Ser which is converted to Ala, Trp which is identified as Gly + Ala, and Thr which is identified as α-aminobutyric acid. Ser-Ala assignments were made after observing gc profiles in which these two amino acids are easily distinguishable. Leu-Ile co-elute on gc and are differentiated by the HI method in which Pth-Ile is regenerated as a mixture of *allo*-Ile and Ile and Pth-Leu is regenerated to Leu. Most residues in the latter portion of each sequencer run were also hydrolyzed so that identifications were made by both gc and liquid chromatography. Amino acids were assigned at each position after observing the rise and fall of specific amino acids in both gc and analyzer profiles at a given, as well as the preceding and following steps (Hermodson *et al.*, 1972). Where possible acids and amides were differentiated by thin-layer chromatography (Capra *et al.*, 1972). Cysteine was identified by liquid scintillation counting of the ^{14}C-labeled derivative.

Results

Heavy Chain NH_2 Terminal Sequence. Uncleaved heavy chains were subjected to 37 cycles of degradation and the sequence obtained is given in Table I. This NH_2 terminal sequence as well as that for heavy chains of other phosphorylcholine binding myeloma proteins have previously been reported (Barstad *et al.*, 1974).

Heavy Chain Cyanogen Bromide Fragments. The partially reduced and alkylated heavy chain was cleaved with cyanogen bromide and then chromatographed on a Sephadem G-100 column (2.5 × 125 cm) in 5 M guanidine–0.2 M NH_4HCO_3 (Figure 1A). The two major protein peaks were pooled as indicated, desalted on Sephadex G-10 columns in 0.1 M NH_3, and lyophilized.

Cn-1 was dissolved in 5 M guanidine–0.05 M Tris (pH 8.2) and fully reduced with 10 mM dithiothreitol for 2 hr at 37°, followed by alkylation with 20 mM [^{14}C]iodoacetamide for 30 min at 37°. Cn-1 was then applied to the same Sephadex G-100 column as described above and eluted under the same conditions. The major peaks of absorbance and radioactivity were pooled as indicated in Figure 1B, desalted, and lyophilized.

Cn-1A was subjected to 23 cycles of degradation and found to be identical with the large C region fragment found by Francis *et al.* (1974) in the MOPC 315 heavy chain and will be described in a subsequent communication. Cn-1B gave ambiguous results in a preliminary sequencer experiment and since the yield in terms of both absorbance and radioactivity was quite low, it was not analyzed further.

Cn-1D was shown to have a sequence identical with that of the heavy chain amino terminus and was therefore designated as the NH_2 terminal fragment. Cn-1C in a preliminary sequencer experiment was found to be a single fragment and was subsequently subjected to 35 cycles of degradation. The sequence of this fragment is given in Table I and its position in the chain can be assigned as extending from residue 84 in the V region to residue 116 (the first residue in the C region) by homology with other heavy chain sequences from both mice and humans

TABLE I: Amino Acid Sequence and Alignment of CNBr Fragments.

	1 10 20
Heavy chain NH_2 terminus[a]	Glu-Val-Lys-Leu-Val-Glu-Ser-Gly-Gly-Gly-Leu-Val-Gln-Pro-Gly-Gly-Ser-Leu-Arg-Leu-Ser-Cys-Ala-Thr-Ser
Cn-1C[a]	Asn-Ala-Leu-Arg-Ala-Glu-Asp-Thr-Ala-Ile-Tyr-Tyr-Cys-Ala-Arg-Asn-Tyr-Tyr-Gly-Ser-Thr-Trp-Tyr-Phe-Asp
Cn′-2sed	Glu-Trp-Val-Arg-Gln-Pro-Pro-Gly-Lys-Arg-Leu-Glu-Trp-Ile-Ala-Ala-Ser-Arg-Asx-Lys-Gly-Asx-Lys-Tyr-Thr
	30 40
Heavy chain NH_2 terminus	Gly-Phe-Thr-Phe-Ser-Asx-Phe-Tyr-Met-Glu-Trp-Val
Cn-1C	Val-Trp-Gly-Ala-Gly-Thr-Thr-Val-Thr-Val
Cn′-2sed	Thr-Glx-Tyr-Ser-Ala-Ser-Val-Lys-Gly-Arg-Phe-Ile-Val-Ser-Arg-Asx-Thr-Ser-Glx-Ser-Ile-Leu-Tyr-Leu-Gln-Hse
	← CPA

Alignment of CNBr Fragments

1		34	37	83	116
heavy chain NH_2 terminus		Met			
			Cn′-2sed	Met	Cn-1C

[a] Repetitive yields from representative sequencer runs of the heavy chain amino terminus and Cn-1C were 94.2 and 95.2% respectively.

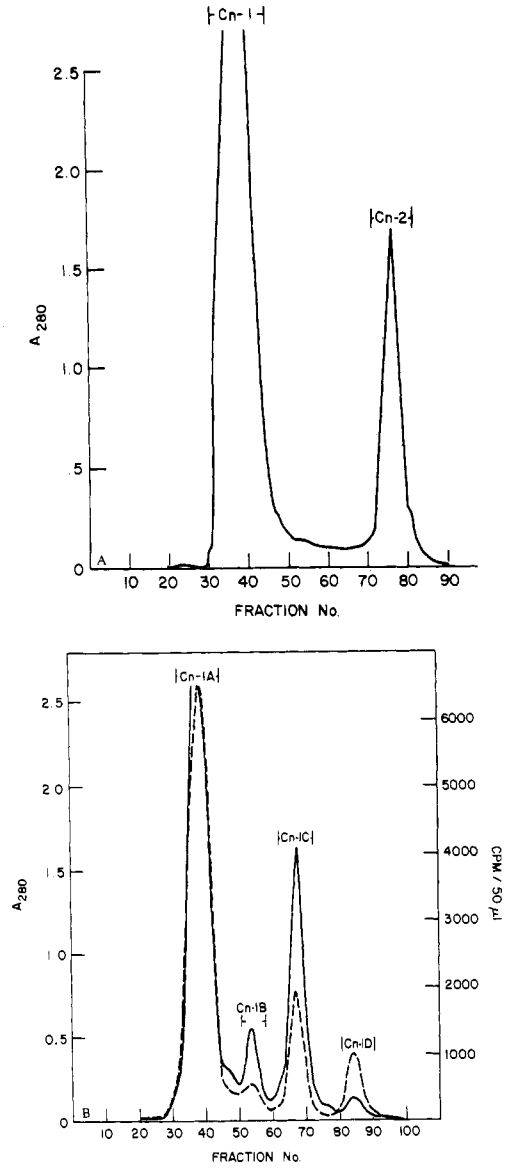

FIGURE 1: (A) Sephadex G-100 gel filtration of CNBr cleaved M603 heavy chain. Columns (125 × 2.5 cm) were equilibrated and eluted with 5 M guanidine·HCl–0.2 M NH_4HCO_3. (B) Gel filtration as in A of Cn-1 following full reduction and alkylation with [^{14}C]iodoacetamide.

(Figure 2). From the presence of the methionine at position 34 in the NH_2 terminal sequence and the location of Cn-1C (residues 84–116) it was obvious that the fragment from residue 34 to 84 would not be disulfide linked to any other fragment since the V region intrachain disulfide is between the half-cysteines at positions 22 and 96. Cn-2 was therefore examined as the likely location of the missing V region fragment. Cn-2 was found by sequence analysis to be a complex mixture of approximately four fragments. It was therefore decided to use pepsin Fab fragments in an attempt to reduce the complexity of this mixture.

Pepsin Fab Cyanogen Bromide Fragments. Fab was cleaved with cyanogen bromide and chromatographed under the same conditions as described for the heavy chain. An elution profile of the cleaved Fab is presented in Figure 3. Pools were made as indicated, desalted, and lyophilized. Cn′-2 from the Fab was found to be a mixture of two fragments by sequence analysis. It was fortuitously found that upon suspension in 0.01 M NH_4OAc (pH 5.0) one of the peptides became largely insoluble. The insoluble peptide was collected by centrifugation after being allowed to stand overnight at room temperature and dissolved in 5% HOAc. The amino acid composition of this fragment (subsequently referred to as Cn′-2sed) is given in Table II along with the number of residues of each amino acid found in subsequent sequencing experiments.

TABLE II: Amino Acid Composition of Cn′-2sed.

Amino Acid	Residues/mol[a]	Obsd
Asp	3.0	3
Thr	3.0	3
Ser	5.8	6
Glu	6.4	6
Pro	1.8	2
Gly	3.1	3
Ala	2.9	3
Val	2.8	3
Ile	2.9	3
Leu	3.1	3
Tyr	3.0	3
Phe	0.9	1
Lys	3.8	4
Arg	4.8	5
Trp	*b*	2
Hse	1.1	1

[a] Values are the average of two separate 20-hr hydrolysates. [b] Not determined.

Approximately 400 nmol of Cn′-2sed was applied to the sequencer and subjected to 48 cycles of degradation. The first 48 residues in this fragment were thus identified by sequencer analysis. An aliquot was then digested with carboxypeptidase A and samples were taken as indicated in Figure 4 and applied to the amino acid analyzer. Amino acids were released in the order homoserine, glutamine, leucine, and tyrosine. As tyrosine was identified at position 48 in the sequencer analysis the overlap is provided and establishes the sequence of the C-terminal four positions as Tyr-Leu-Gln-Hse. The complete sequence of this fragment is given in Table I. The amino terminus of Cn′-2sed overlaps the C-terminal three residues of the intact heavy chain sequencer run (Table I) and places this fragment as extending from position 35 to 83 in the variable region (Figure 2).

Discussion

With the exception of the possible but highly unlikely occurrence of an insertion between positions 83 and 84 we have determined the entire V region sequence of the heavy chain from protein M603. At present the complete variable region sequence of only one other mouse heavy chain, M315, has been reported (Francis *et al.*, 1974), although Bourgois *et al.* (1972) have previously described the first 104 positions of the M173 heavy chain. The presence of methionine residues at positions 34 and 83 in the V_H permitted the isolation of the entire variable region in three fragments following CNBr cleavage.

When the heavy chains from mouse and human proteins are aligned in a manner to maximize homology (Figure 2) it becomes apparent that variations in chain lengths may all be attributable to insertions or deletions in the hypervariable regions (31–35, 50–62, 99–105). The M603 heavy chain has an unusual insertion of two residues following position 58 which is in the second hypervariable region and a second two-residue insertion

		10	20	30
Mouse 603	(α,κ)	E V K L V E S G G G L V Q P G G S L R L S C A T S G F T F S		
Mouse 315	(α,λ2)	D V Q L Q E S G P G L V K P S Q S L S L T C S V T G Y S I T		
Mouse 173	(γ2a,κ)	E V K L L E S G G P L V Q L G G S L K L S C A A S G F D F S		
Human Nie	(γ1,κ,V_{HIII})	Z V Q L V Q S G G G V V Q P G R S L R L S C A A S G F T F S		
Human Ou	(μ,κ,V_{HII})	Z V T L T E S G P A L V K P K Q P L T L T C T F S G F S L S		
Human Eu	(γ1,κ,V_{HI})	Z V Q L V Q S G A E V K K P G S S V K V S C K A S G G T F S		
		40	50	
Mouse 603	(α,κ)	B F Y M E - - W V R Q P P G K R L E W I A A S R B K G B K Y		
Mouse 315	(α,λ2)	S G Y F K N - W I R Q F P G N K L E W L G F I K Y D G S B -		
Mouse 173	(γ2a,κ)	R Y W M S - - W V K Q A P G K G L E W I G E I D P N S S T I		
Human Nie	(γ1,κ,V_{HIII})	R Y T I H - - W V R Q A P G K G L E W V A V M S Y B G B B K		
Human Ou	(μ,$\kappa$$V_{HII}$)	T S R M R V S W I R R P P G K A L E W L A R I - - B B B D K		
Human Eu	(γ1,κ,V_{HI})	R S A I I - - W V R Q A P G Q G L E W M G G I V P M F G P P		
		60	70	80
Mouse 603	(α,κ)	T T Z Y - S A S V K G R F I V S R B T S Z S I L Y L Q M N A		
Mouse 315	(α,λ2)	- - - (Y G)B P S L K N R V S I T R D T S E N Q F F L K L D S		
Mouse 173	(γ2a,κ)	- - N Y - T P S L K D K F I I S R N D A K N T L Y L Q M S K		
Human Nie	(γ1,κ,V_{HIII})	- - H Y - A D S V N G R F T I S R N D S K N T L Y L N M N S		
Human Ou	(μ,κ,V_{HII})	- - F Y W S T S L R T R L S I S K N D S K N Q V V L I M I N		
Human Eu	(γ1,κ,V_{HI})	- - N Y - A Q K F Q E R V T I T A D E S T N T A Y M E L S S		
		90	100	
Mouse 603	(α,κ)	L R A E D T A I Y Y C A R N Y Y G S T W Y - - - - - F D V W		
Mouse 315	(α,λ2)	V T (T Z B) T A T Y Y C A G D N D H L Y - - - - - - - F D Y W		
Mouse 173	(γ2a,κ)	V R S E D T A L Y Y C A R S P Y Y A M		
Human Nie	(γ1,κ,V_{HIII})	L R P Z B T A V Y Y C A R I R D T A M F - - - - - - F A H W		
Human Ou	(μ,κ,V_{HII})	V N P V D T A T Y Y C A R V V N S V M A G Y Y Y Y Y M D V W		
Human Eu	(γ1,κ,V_{HI})	L R S E D T A F Y F C A G G Y G I Y S - - - - - - - P E E Y		
		110		
Mouse 603	(α,κ)	G A G T T V T V		
Mouse 315	(α,λ2)	G Q G T T L T V		
Human Nie	(γ1,κ,V_{HIII})	G Q G T L V T V		
Human Ou	(μ,κ,V_{HII})	G K G T T V T V		
Human Eu	(γ1,κ,V_{HI})	N G G - L V T V		

FIGURE 2: Heavy chain variable region sequence from M603 compared with sequences from mouse proteins M315 (Francis *et al.*, 1974), M173 (Bourgois *et al.*, 1972), and human proteins Nie (Ponstingl *et al.*, 1970), Ou (Shimizu *et al.*, 1971a,b), and Eu (Edelman *et al.*, 1969; Edelman, 1970). Numbering is according to Francis *et al.* (1974) with gaps introduced to maximize homologies.

following position 104 which is in the third hypervariable region. A comparison of the available mouse heavy chain sequences (Figure 2) reveals that 40 positions in the V_H are invariant. If one excludes the hypervariable regions which are thought to determine antigen binding specificity, these 40 positions comprise 44% of the remainder of the variable region. Furthermore, if one considers changes such as Leu-Ile-Val, Gly-Ala, Ser-Thr, or Lys-Arg as those which would not affect the functional homology at a given position the number increases to 59%. A similar comparison including the human heavy chains in Figure 2 reveals 20 invariant positions (22%). This number increases to 41% if the interchanges described above (Leu-Ile-Val, etc.) are considered as homology positions.

The V_H homology between man and mouse suggests that genes coding for these sequences have been highly conserved throughout evolution. This observation is further supported by data of Capra *et al.* (1973) who have described unblocked heavy chain amino-terminal sequences from a number of species and similarly observed this large degree of homology.

At this time it appears as though there are several basic questions in immunology to which sequence analyses will provide essential information. First, primary sequences will be necessary to construct three-dimensional models of crystallized immunoglobulin molecules. In this regard, Padlan *et al.* (1973) have previously reported the structure of the M603 Fab at 4.5-Å resolution. A molecular model at 3.1-Å resolution is currently being constructed from which it can be seen that a wedge-shaped cleft is formed at one end of the molecule by loops consisting of the three hypervariable regions of the heavy chain and two of the hypervariable regions of the light chain. Phosphorylcholine has been demonstrated to bind in this cleft with the predominant specific interactions between the hapten and residues in the hypervariable regions of the heavy chain. The two residues in nearest proximity to the hapten are Tyr-33 and Arg-52 both of which are apparently hydrogen bonded to oxygen atoms of the phosphate moiety of the hapten. Lys-54 is also in close proximity to the phosphate and would therefore help neutralize the negative charge in this portion of the hapten. The choline moiety is involved in van der Waals interactions with Gly-102 and Ser-103 as well as residues in the third hypervariable region of the light chain (Padlan *et al.*, 1974). This predominant interaction of hapten with the heavy chain may explain the high degree of conservation of this amino acid sequence in other phosphorylcholine binding proteins in the mouse (see below).

Second, an exploration of the primary structure of groups of proteins which bind the same hapten will provide information as to the types of variation that may occur while still retaining antigen binding specificity. Of five phosphorylcholine binding proteins (T15, H8, S107, M167, and M603) studied to date all have identical heavy chain sequences through the first hypervariable region with the exception of a single substitution at position 4 in the M167 chain. The light chains from three of the proteins (T15, H8, and S107) have identical sequences through the first hypervariable region. The two other light chains (M167 and M603) have greatly different sequences from each other and the H8–T15 pair (Barstad *et al.*, 1974). It therefore seems as though at least three distinct germ line genes are involved in coding for light chains of phosphorylcholine binding proteins. For this antigen binding specificity, then,

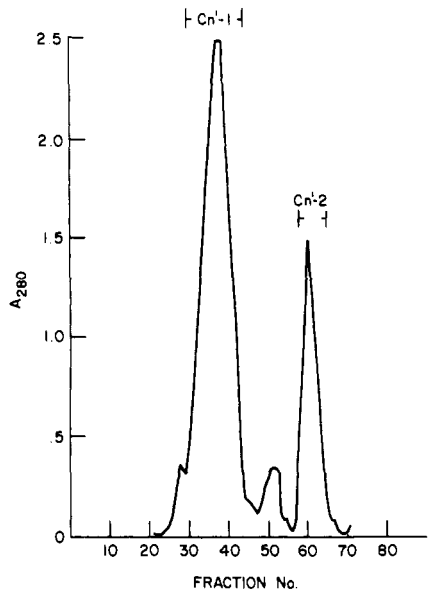

FIGURE 3: Sephadex G-100 gel filtration of Fab following CNBr cleavage. Conditions are the same as in Figure 1A.

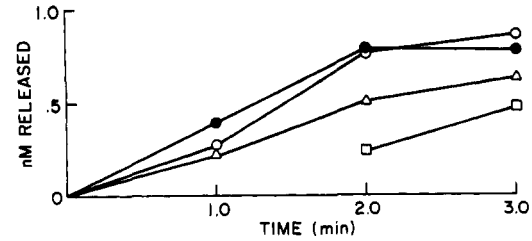

FIGURE 4: Carboxypeptidase A digestion of Cn'-2sed Digestion was performed at room temperature in 0.01 M phosphate buffer (pH 8.0). Amino acids are represented as follows: (●) homoserine; (○) glutamine; (△) leucine; (□) tyrosine.

the heavy chain may play a dominant role. This interpretation is consistent with the X-ray crystallographic studies described above which have demonstrated the primary specific interactions to be between the hapten and the heavy chain. Weigert *et al.* (1970) have described two α-(1→3)-dextran binding myeloma proteins which have identical light chain sequences, but have different heavy chain amino terminal sequences (L. Hood, personal communication). Thus, for α-(1→3)-dextran binding this specific light chain appears to be stringently required for this specificity, although a final conclusion must await V region determinations on the heavy chains of these molecules. We have previously described a group of myeloma proteins which bind β-(1→6)-D-galactan (Rudikoff *et al.*, 1973). Of six proteins studied to date all have identical light chains for the first 23 residues. Four of the heavy chains from these proteins are identical for their first 30 residues and the other two have only one substitution. For this specificity a particular light as well as heavy chain gene may be necessary. This group of proteins should be extremely interesting for an exploration of structure-function relationships in that even though the light and heavy chain sequences are very similar each protein possesses a unique idiotype and has different binding characteristics for various galactan haptens (Jolley *et al.*, 1973). Thus, the predominance of specific interactions between hapten and light chain or heavy chain or a light-heavy combination may depend greatly on the antigen binding specificity required and not follow any general rule in regard to importance of light *vs.* heavy chain in determining specificity.

Finally, Claflin *et al.* (1974) have recently reported the occurrence of antiphosphorylcholine antibodies in strains of mice other than BALB/c which appear to be of restricted heterogeneity and have patterns of binding specificity for phosphorylcholine derivatives similar to antibodies induced in the BALB/c strain. The binding specificity of these induced antibodies is identical with that of the phosphorylcholine binding myeloma proteins T15 and H8 described above. Sequence analysis of the antibodies induced in these various strains may permit an evaluation of the variation of a distinct pair of germ line genes within a given species.

Added in Proof

The electron density map of the M603 heavy chain is consistent with our supposition that no insertion occurs between positions 83 and 84.

Acknowledgment

We are grateful to Dr. Sharon Francis for making her sequence of the MOPC 315 heavy chain available to us prior to publication and for valuable discussions throughout this work. We thank Ms. Mary Heller for excellent technical assitance.

References

Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W., and Hood L. (1974), *Science 183,* 962.

Bourgois, A., Fougereau, M., and De Preval, C. (1972), *Eur. J. Biochem. 24,* 446.

Bridges, S. H., and Little, J. R. (1971), *Biochemistry 10,* 2525.

Capra, J. D., Kehoe, J. M., Williams, R. C., Feizi, T., and Kunkel, H. G. (1972), *Proc. Nat. Acad. Sci. U. S. 69,* 40.

Capra, J. D., Wasserman, R. L., and Kehoe, J. M. (1973), *J. Exp. Med. 138,* 410.

Chesebro, B., and Metzger, H. (1972), *Biochemistry 11,* 766.

Cisar, J., Kabat, E. A., Liao, J., and Potter, M. (1974), *J. Exp. Med. 139,* 159.

Claflin, J. L., Lieberman, R., and Davie, J. M. (1974), *J. Immunol. 112,* 1747.

Edelman, G. M. (1970), *Biochemistry 9,* 3197.

Edelman, G. M., Cunningham, W. E., Gottlieb, P. D., Rutihauser, V., and Waxdal, M. J. (1969), *Proc. Nat. Acad. Sci. U. S. 63,* 78.

Eisen, H. N., Simms, E. S., and Potter, M. (1968), *Biochemistry 7,* 4126.

Francis, S. H., Leslie, G. Q., Hood, L., and Eisen, H. N. (1974), *Proc. Nat. Acad. Sci. U. S. 71,* 1123.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11,* 4493.

Jaffe, B. M., Eisen, H. N., Simms, E. S., and Potter, M. (1969), *J. Immunol.* 103, 872.

Jolley, M. E., Rudikoff, S., Potter, M., and Glaudemans, C. P. J. (1973), *Biochemistry 12,* 3039.

Leon, M. A., and Young, N. M. (1971), *Biochemistry 10,* 1424.

Leon, M. A., Young, N. M., and McIntire, K. R. (1970), *Biochemistry 9,* 1023.

Padlan, E. A., Segal, D. M., Cohen, G. H., Davies, D. R., Rudikoff, S., and Potter, M. (1974), *in* The Immune System: Genes, Receptors, Signals, Proceeding of the 1974 ICN-UCLA Symposium on Molecular Biology, Cox, C. F., Ed , Academic Press, New York and London, in press.

Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S., and Potter, M. (1973), *Nature (London), New Biol. 245,* 165.

Pisano, J. J., Bronzert, T. J., and Brewer, H. B. (1972), *Anal. Biochem* 45, 43.

Ponstingl, H., Schwarz, J., Reichel, W., and Hilschmann, N. (1970), *Hoppe-Seylers Z. Physiol. Chem. 351,* 1591.
Potter, M., and Leon, M. (1968), *Science 162,* 369.
Potter, M., and Lieberman, R. (1970), *J. Exp. Med. 132,* 737.
Potter, M., Mushinski, E. B., and Glaudemans, C. P. J. (1972), *J. Immunol. 108,* 295.
Rudikoff, S., Mushinski, E. B., Potter, M., Glaudemans, C. P. J., and Jolley, M. E. (1973), *J. Exp. Med. 138,* 1095.
Rudikoff, S., Potter, M., Segal, D. M., Padlan, E. A., and Davies, D. R. (1972), *Proc. Nat. Acad. Sci. U. S. 69,* 3689.
Sher, A., Lord, E., and Cohn, M. (1971), *J. Immunol. 107,* 1226.
Shimizu, A., Kohler, H., Paul, C., Shinoda, T., and Putnam, F. W. (1971a), *Science 173,* 629.
Shimizu, A., Putnam, F. W., Paul, C., Clamp, J. R., and Johnson, I. (1971b), *Nature (London), New Biol. 231,* 73.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10,* 4912.
Weigert, M., Cesari, M., Yonkovich, S. J., and Cohn, M. (1970), *Nature (London) 228,* 1045.

Characterization of Concanavalin A Sugar Binding Site by ^{19}F Nuclear Magnetic Resonance†

Gerald M. Alter and James A. Magnuson*

ABSTRACT: The binding of *N*-trifluoroacetyl-α- and β-D-glucosamine to dimeric and tetrameric concanavalin A was studied using ^{19}F nuclear magnetic resonance. Competition of methyl α-D-mannopyranoside with the trifluoro sugar indicates that the sugar binds to the carbohydrate binding site of concanavalin A. A great deal of rotational freedom of the trifluoroacetyl residue was observed for bound *N*-trifluoroacetyl-D-glucosamine anomers in complexes with two protein aggregation states. No chemical shift in the ^{19}F resonance of the probe sugar was observed upon its binding to concanavalin A. These results indicate that binding of the trifluoro sugar in the protein hydrophobic pocket previously identified by X-ray crystallographic analysis (Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1972), *Proc. Nat. Acad. Sci. U. S. 69,* 2580; Hardman, K. D., and Ainsworth, C. F. (1972), *Biochemistry 11,* 4910) is unlikely. The ^{19}F resonance is broadened noticeably when the sugar binds to concanavalin A containing Mn^{2+} in the transition metal site. This broadening was used to calculate the distance separating the fluorine nuclei and the Mn^{2+} ion. The mean distance between the fluorine nuclei of the bound sugar and the concanavalin A transition metal site was found to be 12 and 14 Å for the α and β anomer, respectively. This is in agreement with carbon magnetic resonance results independently obtained (Brewer, C. F., Sternlicht, H., Marcus, D. M., and Grollman, A. P. (1973), *Proc. Nat. Acad. Sci. U. S. 70,* 1007; Villafranca, J. J., and Viola, R. E. (1974), *Arch. Biochem. Biophys. 160,* 465). It strongly implies that the carbohydrate binding site in solution is not the hydrophobic pocket identified by X-ray crystallographic analysis.

Concanavalin A (Con A[1]) is a protein isolated from the jack bean (*Canavalia ensiformis*) (Sumner and Howell, 1936). It is one of a class of plant proteins called lectins. Lectins, including Con A, have the ability to bind to cell surfaces and for that reason have become widely used in exploring the structure and dynamics of cell surfaces (Sharon and Lis, 1972; Burger, 1973; Inbar *et al.*, 1971). Con A agglutinates embryonic tissue cells (Moscona, 1971) and various neoplastic cells in tissue culture (Inbar and Sachs, 1969). Adult and normal cells are not agglutinated and generally have a lower binding affinity for Con A. Upon binding, Con A often causes a change in the physiological functioning of its target cell (Cuatrecasas, 1973; Powell and Leon, 1970). Burger and Noonan (1970) have demonstrated that normal growth may be restored to virally transformed cells by treatment with trypsinized Con A.

Association of Con A with cell surfaces occurs primarily by virtue of its ability to bind certain carbohydrates on the cell surface (Wray and Waldborg, 1971; Allen *et al.*, 1972). Carbohydrates with the minimum structural requirements for binding to Con A contain residues with the D-arabinopyranoside configuration at the C-3, C-4, and C-6 positions (Goldstein *et al.*, 1965, 1973). There is one monosaccharide binding site per 25,500 molecular weight subunit of Con A (Yariv *et al.*, 1968).

For monosaccharide binding activity, one transition metal and one calcium ion must be bound to each Con A subunit (Kalb and Levitzki, 1968). The transition metal requirement is satisfied by several metal ions including diamagnetic zinc and paramagnetic manganese ions. The calcium requirement, however, is more specific. Only cadmium and strontium substitution yields an active protein (Shoham *et al.*, 1973).

Sedimentation equilibrium studies indicate Con A exists as a 110,000 molecular weight tetramer at pH 7.0 and higher and

† From the Graduate Program in Biochemistry and Biophysics, Department of Chemistry, Washington State University, Pullman, Washington 99163. *Received March 27, 1974.* This research was supported in part by a grant from the Washington State University Research Committee and in part by the U. S. Public Health Service Grant CA 14496-01. A preliminary account of this work was delivered at the Pacific Slope Biochemical Conference, Vancouver, British Columbia, Canada, Aug 1973.

[1] Abbreviations used are: Con A, concanavalin A; NTFAGlcn, *N*-trifluoracetyl-D-glucosamine; NphGlcp, *p*-nitrophenyl α-D-glucopyranoside; αMeMan, methyl α-D-mannopyranoside; Mn-Con A, manganese enriched concanavalin A; Zn-Con A, zinc enriched concanavalin A; IphGalp, *O*-iodophenyl β-D-galactopyranoside; IphGlcp, *o*-iodophenyl β-D-glucopyranoside.