Proc. Natl. Acad. Sci. USA
Vol. 73, No. 6, pp. 2109–2112, June 1976
Immunology

Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins

(amino acid sequence/hypervariable regions/insertions–deletions/diversity)

STUART RUDIKOFF AND MICHAEL POTTER

Laboratory of Cell Biology, National Cancer Institute, Bethesda, Maryland 20014

Communicated by Elvin A. Kabat, March 24, 1976

ABSTRACT **The entire sequences of the heavy chain variable regions of M167 and TEPC 15 (phosphorylcholine-binding myeloma proteins of BALB/c origin) have been determined. These sequences are compared with the phosphorylcholine-binding protein M603. T15 differs from M603 at four positions, all of which are located in antigen-binding complementarity regions. M167, in addition to having differences in the complementarity regions, also has five substitutions in the conserved framework portion of the variable region when compared to T15 and M603. Each of the three proteins has a different length in the third complementarity region. It is unlikely that complementarity regions of different lengths associated with similar framework regions could be generated by proposed mechanisms of somatic mutation which are generally limited to point mutations. It appears more likely that these products are directly encoded by different structural germ line genes.**

The mechanism(s) of generation of immunoglobulin diversity has long been a subject of major controversy. Attempts have been made to approach this problem by primary sequence analysis (1–4), studies of the inheritance of serological markers (5–7), and more recently nucleic acid hybridization (8–14). The mouse plasmacytoma system provides a unique opportunity to observe the variation in primary structure among proteins with similar antigen binding specificity. Since the proteins are derived from a single haplotype, an opportunity may be provided to assess the number of structural genes compatible with the observed protein amino acid sequences. This approach is made possible by the previous reports of groups of myeloma proteins with binding activity to chemically defined haptenic determinants (15).

We have previously reported the amino terminal sequences of the light and heavy chains from five phosphorylcholine*-binding myeloma proteins (T15, H8, S107, M603, and M167) and more recently the entire heavy chain variable region from one of these, M603 (16, 17). Three of the proteins, T15, H8, and S107, have identical amino terminal sequences in their light and heavy chains and indistinguishable antigenic determinants of the variable region (idiotypes) (18). This group of proteins also has the same binding specificity for phosphorylcholine and various analogs (19, 20). The other two proteins, M603 and M167, have amino terminal light chain sequences that are quite different from each other and from the T15, H8, S107 group. M603 and M167 also have unique idiotypes (18) as well as different specificities for phosphorylcholine and several analogs (19, 20). In the present communication we report the sequence of the heavy chain variable region of the M167 protein and compare its structure with that of two other heavy chains from antiphosphorylcholine proteins.

MATERIALS AND METHODS

Plasmacytoma MOPC167 (IgA/κ) with binding specificity for phosphorylcholine has previously been described by Potter and Leon (21) and Leon and Young (19). The protein was purified by immunoadsorption and heavy and light chains separated as previously reported (17).

Peptide Fragments. Cyanogen bromide fragments were generated and purified in the same manner as described for the M603 heavy chain (17). Where necessary CNBr fragments were digested with trypsin (1:100) in 0.2 M ammonium bicarbonate and the resulting tryptic peptides were purified by high voltage paper electrophoresis.

Sequence Determination. Automated Edman degradations were performed on a Beckman 890C sequencer. Thiazolinones obtained after each cycle were converted to phenylthiohydantoins and identified as previously reported (17, 22). All positions were initially analyzed by gas chromatography. Amino acids not conclusively identified by gas chromatography were hydrolyzed to their free amino acids and subjected to amino acid analysis. Thin-layer chromatography was employed primarily to differentiate acidic amino acids from their amides. In addition, all positions in the latter part of sequencer runs were hydrolyzed and subjected to amino acid analyses. Acids and amides not differentiated during sequencer analysis were subsequently determined by several methods including digestion with leucine aminopeptidase or carboxypeptidase A or mobility on high voltage paper electrophoresis. Certain tryptic peptides containing homoserine at the C-terminal position were coupled to solid phase sequencing resin by the method of Horn and Laursen (23) and degraded on a Sequemat model 12 solid phase sequencer.

RESULTS

The variable region sequence of the M603 heavy chain has previously been reported (17). Acid and amide residues not previously differentiated have been determined and are included in Table 1.

CNBr fragments

M167 heavy chains were cleaved with CNBr and chromatographed on a Sephadex G-100 column in 5 M guanidine–0.2 M ammonium bicarbonate (Fig. 1A). The resulting peaks were pooled as indicated, desalted on Sephadex G-10 columns in 0.1 M ammonia and lyophilized. Cn-1 was then fully reduced in 5 M guanidine–0.05 M Tris at pH 8.2, alkylated with [^{14}C]-iodoacetamide and applied to the same Sephadex G-100 column. The elution pattern is given in Fig. 1B. The profiles in Fig. 1A and B are identical to that of M603 and result from the presence of methionine residues at positions 34 and 83 in the variable region of both proteins.

Sequence of CNBr fragments

All CNBr fragments were subjected to at least two sequencer runs. Cn-1A has previously been reported as a large fragment

* The Commission on Biochemical Nomenclature recommends the name *phosphocholine* for this compound.

Table 1. Sequences of heavy chain variable regions from phosphorylcholine-binding proteins

	1	10	20	30 $HV_I{}^b$ →	40
T15[a]	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 D F Y M E W V R Q P P				
M603	—				
M167	—V—				—T—

	50 ← HV_{II}	A B 60 →	70	80
T15	G K R L E W I A A S R N K A N D Y T T E Y S A S V K G R F I V S R D T S Q S I L Y L			
M603	—G—K—			
M167	—S—H—R—			—V—

	90	100 ← HV_{III} →	110
T15	Q M N A L R A E D T A I Y Y C A R D — Y Y G S S Y — W Y F D V W G A G T T V T V		
M603	—N — —T — —		
M167	—T—T—A D—(D)—F G—		

[a] Details of the T15 sequence are being published separately (S. Rudikoff, M. Potter, P. Barstad, and L. Hood, manuscript in preparation).
[b] HV_I, HV_{II}, HV_{III} correspond to the hypervariable or complementarity regions discussed in the *text*. Positions for deletions in HV_{III} were selected in an attempt to maintain maximum homology.

from the constant region (17, 24). Cn-1C is the fragment beginning at position 83 in the variable region and extending approximately to residue 160. This fragment was sequenced to position 116, the beginning of the constant region. Cn-1D was found to be the amino terminal fragment (residues 1–34). The amino terminal sequence of this protein has previously been reported (16).

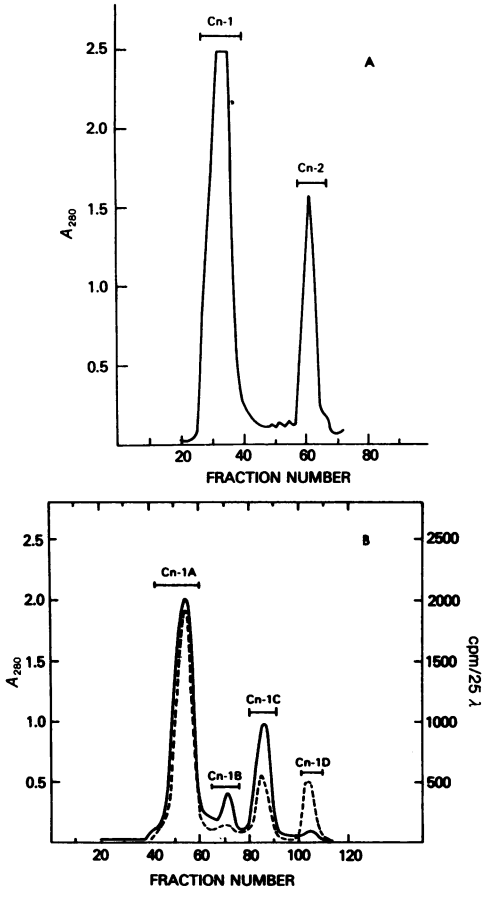

FIG. 1. (*A*) Elution profile of CNBr-cleaved M167 heavy chain on Sephadex G-100 in 5 M guanidine–0.2 M ammonium bicarbonate. (*B*) Elution profile of Cn-1 after full reduction and alkylation with [^{14}C]iodoacetamide. Chromatography conditions are the same as in *A*. A_{280}——; cpm – – –.

The remaining variable region peptide (residues 35–83) was purified from Cn-2 or from pepsin Fab fragments by virtue of its insolubility in the same manner as previously reported for M603 (17). The C-terminal 11 residues from this fragment were confirmed by digesting the intact fragment with trypsin and specifically coupling the C-terminal homoserine containing peptide to triethylenetetramine resin (23). The peptide was then sequenced on a solid phase instrument. The entire variable region sequence of M167 is presented in Table 1.

DISCUSSION

The presence of methionine residues at positions 34 and 83 of the heavy chain variable region in all mouse phosphorylcholine-binding proteins has facilitated their sequence determination by the procedures employed for M603 (17). These proteins provide a unique opportunity to evaluate primary structural differences among a group of antibodies demonstrating activity for the same chemically defined hapten.

Each of the three proteins (M603, T15, and M167) discussed has previously been shown to possess a light chain from a different kappa subgroup (3) and also individual antigenic determinants (idiotypes) (18). However, in any discussion of genetic mechanisms involved in the ultimate expression of the observed structures it should be kept in mind that the T15 idiotype has been found in eight independently induced phosphorylcholine-binding myeloma proteins whereas the others have occurred only once (18, 25). Also, this idiotype is present in the serum of all normal conventionally raised BALB/c mice in our colony (26) and is the predominant idiotype detected after immunization with phosphorylcholine containing antigens (20, 27–30). Thus, genes controlling the T15 structure or structures very similar to T15 must be maintained in the germ line.

At present, two major hypotheses being entertained to explain the generation of antibody diversity are the germ line theory and the somatic mutation theory. According to the first, the information coding for all variable region sequences is contained in the germ line. The somatic mutation theory, on the

other hand, proposes the existence of a relatively few germ line genes and attributes the subsequent generation of diversity to mutation occurring in somatic cells. Clearly, these two hypotheses need not be mutually exclusive.

In the comparison of the heavy chain sequences in Table 1, T15 has been selected as the prototype because of the predominance of this idiotype in normal and induced phosphorylcholine-binding antibodies in mice. When M603 is compared with T15, four amino acid interchanges are seen: two in the second and two in the third hypervariable regions. In the following discussion hypervariable regions correspond to the three complementarity regions in heavy chains described by Kabat and Wu (31) and subsequently shown by crystallographic analysis to be those regions that form the surface of the combining site contributed by the heavy chain (32, 33). Only the Lys-Asp interchange at position 57 would require more than a single base mutation to explain the observed amino acid differences. M167 differs from T15 at 13 positions, eight of which are in hypervariable regions (three in HV_{II} and five in HV_{III}) and five of which are scattered throughout the framework. Even though many of these substitutions result in structurally very different amino acids, e.g., Arg for Thr in the second hypervariable region, all of the interchanges, with the exception of the tentative Ser–Asp exchange in the third hypervariable region, can also be explained by single base mutations.

Possibly the most striking observation when comparing these sequences is the size differences in the third hypervariable region. M167 can be seen to be three residues longer than M603 and two residues longer than T15. On the basis of the observed amino acid interchanges it is impossible to decide whether T15 and M603 represent different germ line genes or are somatic derivatives. Studies on induced antibodies against phosphorylcholine from A/J mice have revealed the presence of the T15, M603, and M167 idiotypes in approximately equimolar ratios (S. Rudikoff and L. Claflin, manuscript in preparation). This observation suggests that the M603 and T15 sequences are encoded in the germ line and are the products of different structural genes. The 13 amino acid differences observed between M167 and T15 as well as the size difference indicates the presence of an additional germ line gene coding for this heavy chain.

Capra and Kehoe (34) have reported the heavy chain sequences from two human myeloma proteins with antigamma globulin activity. Eight differences were noted between the two with seven of the differences occurring in regions not associated with antigen binding. It is difficult to speculate as to whether these differences represent the expression of alleles due to the genetic polymorphism present in humans or genetic events in the ontogeny of the lymphocyte population. No size difference was observed between these two proteins.

Assuming these three very similar sequences represent at least two and possibly three different germ line gene products it is interesting to speculate on their origin. A plausible explanation is that selective pressure on the T15 prototype has resulted in the duplication of this gene many times in the germ line. Since there is no *a priori* reason to assume that mutations do not occur in immunoglobulin germ line genes as in other systems, the M603 and M167 genes could have resulted from mutation of the T15 prototypes, yet the T15 gene would be continually selected for and expanded permitting its maintenance while during the same evolutionary period there would be an accumulation of other germ line genes (M603 and M167) which produce phosphorylcholine reactive proteins.

The size differences among such closely related structures provide a unique finding which requires the introduction of an additional mechanism which would permit the insertion or deletion of small amounts of nucleic acid to account for these observed differences. If one examines heavy chain sequences from both mice and humans (see ref. 17) it is apparent that most size differences occur in the third hypervariable region although differences have been noted in the other two also. Deletion–insertion mutations in complementarity regions may in some instances be of selective advantage to a species in that size changes would greatly affect the shape of the combining site and could thus potentially create a new group of binding specificities which would further be modified by single amino acid changes. Enzymes such as terminal transferase (35) could conceivably be involved in the origin of deletion–insertion type mutations. We are currently exploring the primary structure of groups of galactan- (22) and levan- (36) binding proteins to determine if the pattern of variation observed in phosphorylcholine-binding proteins is unique to this group or is found among other closely related proteins. It thus appears that the three heavy chain sequences compared represent the products of at least two different structural germ line genes. If several germ line genes exist for phosphorylcholine and a comparable situation exists for other antigen binding specificities it seems likely that the number of genes encoded in the germ line will be quite large. This does not preclude in any way the occurrence of somatic mutation during the expression of these genes.

1. Weigert, M., Cesari, M., Yonkovich, S. J. & Cohn, M. (1970) *Nature* **228**, 1045–1047.
2. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12**, 760–771.
3. Hood, L., McKean, D., Farnsworth, V. & Potter, M. (1973) *Biochemistry* **12**, 741–748.
4. Capra, J. D., Wasserman, R. L. & Kehoe, J. M. (1973) *J. Exp. Med.* **138**, 410–427.
5. Eichmann, K. & Kindt, T. J. (1971) *J. Exp. Med.* **134**, 532–552.
6. Pawlak, L. L., Mushinski, E. B., Nisonoff, A. & Potter, M. (1973) *J. Exp. Med.* **137**, 22–31.
7. Eichmann, K. & Berek, C. (1973) *Eur. J. Immunol.* **3**, 599–601.
8. Delovitch, T. & Baglioni, C. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 739–745.
9. Tonegawa, S., Bernardini, A., Weimann, B. J. & Steinberg, C. (1974) *FEBS Lett.* **40**, 92–96.
10. Faust, C. H., Diggelman, H. & Mach, B. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2491–2495.
11. Stavnezer, J., Huang, R. C. C., Stavnezer, E. & Bishop, J. M. (1974) *J. Mol. Biol.* **88**, 43–63.
12. Honjo, T., Packman, S., Swan, D., Nav, M. & Leder, P. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3659–3663.
13. Premkumar, E., Shoyab, M. & Williamson, A. R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 99–103.
14. Rabbitts, T. H. & Milstein, C. (1975) *Eur. J. Biochem.* **52**, 125–133.
15. Potter, M. (1972) *Physiol. Rev.* **52**, 632–719.
16. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183**, 962–964.
17. Rudikoff, S. & Potter, M. (1974) *Biochemistry* **13**, 4033–4038.
18. Potter, M. & Lieberman, R. (1970) *J. Exp. Med.* **132**, 737–751.
19. Leon, M. A. & Young, N. M. (1971) *Biochemistry* **10**, 1424–1428.
20. Claflin, J. L., Lieberman, R. & Davie, J. M. (1974) *J. Immunol.* **112**, 1747–1756.
21. Potter, M. & Leon, M. A. (1968) *Science* **162**, 369–371.
22. Rudikoff, S., Mushinski, E. B., Potter, M., Glaudemans, C. P. J. & Jolley, M. E. (1973) *J. Exp. Med.* **138**, 1095–1106.
23. Horn, M. J. & Laursen, R. A. (1973) *FEBS Lett.* **36**, 285–288.
24. Francis, S. H., Leslie, G. Q., Hood, L. & Eisen, H. N. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1123–1127.
25. Sher, A., Lord, E. & Cohn, M. (1971) *J. Immunol.* **107**, 1226–1234.

26. Lieberman, R., Potter, M., Mushinski, E. B., Humphrey, W. & Rudikoff, S. (1974) *J. Exp. Med.* **139**, 983–1001.
27. Cosenza, H. & Kohler, H. (1972) *Science* **176**, 1027–1029.
28. Sher, A. & Cohn, M. (1972) *Eur. J. Immunol.* **2**, 326–331.
29. Claflin, J. L. & Davie, J. M. (1974) *J. Exp. Med.* **140**, 673–686.
30. Gearhart, P. J., Sigal, N. H. & Klinman, N. R. (1975) *J. Exp. Med.* **141**, 56–71.
31. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* **190**, 382–391.
32. Amzel, L. M., Poljak, R. J., Saul, F., Varga, J. M. & Richards, F. F. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1427–1430.
33. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4298–4302.
34. Capra, J. D. & Kehoe, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4032–4036.
35. Bollum, F. J. (1974) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), Vol. 10., pp. 145–171.
36. Lieberman, R., Potter, M., Humphrey, W., Mushinski, E. B. & Vrana, M. (1975) *J. Exp. Med.* **142**, 106–119.