Masuda, T., Kuribayashi, K., and Hanaoka, M. (1969), *J. Immunol. 102*, 1156.
McFarlane, A. S. (1958), *Nature (London) 182*, 53.
Meinke, G. A. C., and Spiegelberg, H. L. (1971), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 468.
Mestecky, J., Kulhavy, R., and Kraus, F. W. (1971a), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 468.
Mestecky, J., Zikan, J., Butler, W. T. (1971b), *Science 171*, 1163.
Miller, F., and Metzger, H. (1965), *J. Biol. Chem. 240*, 3325.
Miller, F., and Metzger, H. (1966), *J. Biol. Chem. 241*, 1732.
Mole, I. E., Jackson, S. A., Porter, R. R., and Wilkinson, J. M. (1971), *Biochem. J. 124*, 301.
Morrison, S. L., and Koshland, M. E. (1972), *Proc. Nat. Acad. Sci. U. S. 69*, 124.
O'Daly, J. A., and Cebra, J. J. (1971), *Biochemistry 10*, 3843.
Porter, R. R. (1959), *Biochem. J. 73*, 119.
Prahl, J. W., Mandy, W. J., and Todd, C. W. (1969), *Biochemistry 8*, 4935.
Reisfeld, R. A., Dray, S., and Nisonoff, A. (1965), *Immunochemistry 2*, 155.
Shuster, J. (1971), *Immunochemistry 8*, 405.
Smyth, D. S., and Utsumi, S. (1967), *Nature (London) 216*, 332.
Spring, S. B., Nisonoff, A., and Dray, S. (1970), *J. Immunol. 105*, 653.
Steinberg, A. G. (1969), *Annu. Rev. Genet. 3*, 25.
Stewart, M. W. (1971), *Biochim. Biophys. Acta 236*, 440.
Todd, C. W. (1963), *Biochem. Biophys. Res. Commun. 11*, 170.
Tomasi, T. B., Jr., and Bienenstock, J. (1968), *Advan. Immunol. 9*, 1.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244*, 4406.
Williams, D. E., and Reisfeld, R. A. (1964), *Ann. N. Y. Acad. Sci. 121*, 373.
Wofsy, L., and Burr, B. (1969), *J. Immunol. 103*, 380.
Zacharius, R. M., Zell, T. E., Morrison, J. H., and Woodlock, J. J. (1969), *Anal. Biochem. 30*, 148.

# Mouse Immunoglobulin Chains. A Survey of the Amino-Terminal Sequences of κ Chains†

Leroy Hood,*‡ David McKean,§ Vincent Farnsworth,‡ and Michael Potter||

ABSTRACT: Thirty-one immunoglobulin κ chains from the highly inbred BALB/c mouse have been examined by automatic sequence analysis and compared with thirteen additional proteins from the literature. These 44 proteins represent at least 32 different sequences over their amino-terminal 23 residues. Although there are nine cases in which two or more κ chains have identical sequences, a majority of the proteins clearly differ from one another by multiple nucleotide substitutions. There appears to be a great deal more sequence diversity in the myeloma κ chains of the BALB/c mouse than in the myeloma κ chains of man. A genealogic tree has been constructed from certain of these sequences to illustrate the genetic events required for any mechanism of antibody diversity. This analysis suggests that the mouse κ chains are quite distinct from their human counterparts and that a large number of germ line variable genes are present in the κ family of the BALB/c genome.

The immune system appears to represent an ideal model system for studying the genetic and evolutionary mechanisms of information storage in higher organisms (Hood and Prahl, 1971). The basic structure of immunoglobulins is now largely understood due to the sequence analysis of homogeneous proteins derived from plasma cell tumors (Potter, 1967). The basic subunit for immunoglobulin molecules is composed of two identical light and two identical heavy polypeptide chains generally covalently linked with disulfide bonds (Edelman and Gall, 1969). Each polypeptide chain is made of an amino-terminal variable (V) region, and a carboxy-terminal constant (C) region (Cohen and Milstein, 1967; Smith *et al.*, 1971).[1] The V regions exhibit an enormous diversity of amino acid sequences presumably related to antibody specificity, while each C region has a single sequence for chains of a given class except for genetic polymorphisms. Structural and genetic studies indicate that the synthesis of immunoglobulin chains is regulated by at least three unlinked families of genes containing κ, λ, and heavy chains (see Gally and Edelman, 1971). Since different V region sequences can be associated with the same C region sequence, it seems that two genes (a V and a C) may exist for each immunoglobulin polypeptide chain. A special joining mechanism to unite V and C genes at the DNA level has been postulated (see Hood, 1972).

The availability of large numbers of myeloma proteins has permitted us to ask questions about the genetic and evolu-

† From the Division of Biology, California Institute of Technology, Pasadena, California 91109. *Received September 12, 1972.* Work supported in part by Grant GB 27605 from the National Science Foundation and GM 06965 from the National Institutes of Health. A preliminary report of this material has been published [Hood, L., McKean, D., and Potter, M. (1970), *Science 170*, 1207].

‡ Division of Biology, California Institute of Technology, Pasadena, Calif. 91109.

§ Department of Biology, Johns Hopkins University, Baltimore, Md. Present address: Department of Medical Genetics, University of Wisconsin, Madison, Wis. 53706.

|| National Career Institute, National Institutes of Health, Bethesda, Md. 20014.

[1] The abbreviations used for immunoglobulin chains are those proposed by the World Health Organization: see *Bull. W. H. O. 41*, 975 (1969).

TABLE I: Systems for the Identification of PTH Derivatives.

| | Gas Chromatography | HCl Hydrolysis | HI Hydrolysis | Thin-Layer Chromatography | [$^{14}$C]-CM-Cys |
|---|---|---|---|---|---|
| PTH-Ala | + | + | + | − | − |
| PTH-Val | + | + | + | − | − |
| PTH-Leu | +[a] | +[e] | + | − | − |
| PTH-Ile | +[a] | +[e,f] | +[f] | − | − |
| PTH-Met | + | + | − | − | − |
| PTH-Trp | + | − | − | − | − |
| PTH-Tyr | + | + | + | − | − |
| PTH-Phe | + | + | + | − | − |
| PTH-CM-Cys | +[b] | − | +[h] | − | + |
| PTH-Pro | +[c] | + | + | + | − |
| PTH-Glu | −[d] | + | + | + | − |
| PTH-Gln | −[d] | +[g] | +[g] | + | − |
| PTH-Asp | −[d] | + | + | + | − |
| PTH-Asn | −[d] | +[g] | +[g] | + | − |
| PTH-Arg | − | + | + | − | − |
| PTH-Lys | −[d] | + | + | + | − |
| PTH-His | −[d] | + | + | − | − |
| PTH-Ser | +[b] | − | +[h] | + | − |
| PTH-Thr | +[c] | − | +[i] | + | − |
| PTH-Gly | +[c] | + | + | + | − |

[a] Leu and Ile cochromatograph on gas chromatography. [b] CM-Cys and Ser cochromatograph on gas chromatography. [c] Pro and Thr cochromatograph. The dehydroproduct of Thr cochromatographs with Gly. [d] These residues can be identified on gas chromatograph with freshly packed columns or by silylation. [e] On paper electrophoresis Leu and Ile are difficult to distinguish from the internal standard Nle. [f] In part convert to *a*Ile by acid hydrolysis. [g] Amides are converted to free acids. [h] Converted to Ala. [i] Converted to $\alpha$-aminobutyric acid.

tionary mechanisms for information storage (V region diversity) in the immune system. The rationale for this approach has been to determine the amino acid sequences of many V regions and to ask whether or not patterns of relationship emerge from comparisons either at the protein or nucleotide level.

We chose to approach this search for patterns through an analysis of light chains from myeloma tumors of the BALB/c mouse for a number of reasons. First, these tumors can be induced at will with intraperitoneal injections of mineral oil (Potter, 1967). Myeloma tumors can be frozen for long-term storage so that a large bank of tumors is available for examination. Second, the primary tumors can readily be transmitted to secondary hosts for the production of large quantities of protein. Finally, and most important, the BALB/c mouse is highly inbred and accordingly genetic polymorphisms which might obscure sequence patterns pertinent to genetic mechanisms of antibody diversity are at least minimized, if not eliminated.

We were interested in two questions. First, what is the nature and extent of amino acid sequence diversity in the $\kappa$ chains from the inbred BALB/c mouse? Second, are the patterns of diversity exhibited in BALB/c $\kappa$ chains comparable to those of the more throughly studied $\kappa$ chains of man? This paper presents an analysis of the amino-terminal portion of 31 different BALB/c $\kappa$ chains and compares them with 13 additional chains from the literature. The second and third papers in this series present more complete sequence data on three mouse $\kappa$ chains which are nearly identical for thier amino-terminal 23 residues.

## Materials and Methods

*Protein Isolation.* Bence-Jones proteins (mouse urinary light chains) were isolated from the urine of female BALB/c mice bearing the corresponding plasma cell tumor. The urine was clarified by centrifugation and exhaustively dialyzed against 0.01 M ammonium bicarbonate. If these preparations were significantly contaminated by mouse urinary protein as determined by agar electrophoresis (Potter and Kuff, 1964), ion-exchange chromatography was carried out on DEAE-A25 as described by Potter (1967). Purification of the intact immunoglobulins and their separation into light and heavy chains has been described by Potter (1967).

The IgG, IgH, IgF, and Bence-Jones proteins were selected because of their availability, whereas the IgA proteins were selected because each had antibody-like activity.

*Chemicals.* The sequenator solvents and reagents were purchased from Beckman Instrument Co. Dithiothreitol (Cyclo) (0.001 M) was added to the butyl chloride. PTH[2]-amino acids were purchased from Mann Research Laboratories. [$^{14}$C]Iodoacetamide was purchased from New England Nuclear Corp. The benzene and glacial acetic acid used for thin-layer chromatography were reagent grade.

*Complete Reduction and Alkylation.* Light chains were reduced and alkylated as previously described (Grant and Hood, 1971).

*Automatic Sequenator Analyses.* Native or alkylated proteins were placed in the Beckman Model 890 protein sequencer in 200–300 $\mu$g of 100% trifluoroacetic acid (Pierce) and examined using a program similar to that described by Edman and Begg (1967). The samples were converted and extracted as previously described by Edman and Begg (1967).

*PTH Sample Analysis.* Five systems for the identification of various PTH derivatives are summarized in Table I. Unambiguous identifications are indicated by +'s without superscripts. Most residues were identified by two or more systems.

Generally $^1/_{25}$th of the sample was submitted to gas chromatography (Hewlett-Packard Model 7600 gas chromatograph with an automatic sample injection system) according to Pisano and Bronzert (1969). This procedure permits the identification of 13 PTH derivatives, although certain pairs are indistinguishable (Table I).

Samples (*ca.* $^3/_{25}$ths) not identified on gas chromatography were run on thin-layer chromatography (tlc). Thin-layer chromatography was carried out on Quanta/Gram plates (Q1F) 20 × 20 cm which had been dried immediately before use in a 100° oven for 20 min. Samples (5-$\mu$g) were spotted 1.5 cm from the bottom of the plate and the spots were dried. This plate was put into a sealed, preequilibrated chromatographic tank lined with Whatman No. 3MM paper which had been soaked with solvent [9:1 benzene–acetic acid (glacial) mixture]. Solvent front was allowed to run to within 1 cm of the top and then the plate was removed and dried for 10 min in an oven at 80°. The PTH-amino acids appeared as dark spots against the fluorescent background under short-wave uv light. This method separates Glu from Gln, Asp from Asn,

[2] Abbreviations used are: PTH, phenylthiohydantoin; dansyl, 1-dimethylaminonaphthalene-5-sulfonyl.

TABLE II: Repetitive Yields for Automatic Sequence Analysis.[a]

| Protein | Yield at Step 2[b] | Repetitive Yield (%) | Protein | Yield at Step 2[b] | Repetitive Yield (%) |
|---|---|---|---|---|---|
| T124 | 281 | 93.6 | M870[d] | 90 | 92.0 |
| M63(1)[c] | 310 | 91.1 | H8 | 88 | 97.5 |
| M63(2) | 251 | 96.1 | M843[d] | 120 | 95.3 |
| B61 | 151 | 91.5 | M47[d] | 94 | 94.5 |
| B32 | 126 | 91.1 | M674[d] | 130 | 89.0 |
| M35 | 229 | 95.5 | M773 | 110 | 94.9 |
| M460 | 209 | 95.5 | M265 | 140 | 93.7 |
| M321 | 240 | 94.5 | M167[d] | 165 | 96.3 |
| T173 | 250 | 94.5 | T15 | 81 | 94.6 |
| M461 | 93 | 94.7 | M23[d] | 85 | 93.6 |
| H5 | 245 | 95.5 | M31[d] | 86 | 93.7 |
| M173[d] | 142 | 93.4 | M467[d] | 90 | 92.6 |
| M384[d] | 120 | 90.4 | M603[d] | 90 | 93.1 |
| LPC-1[d] | 59 | 95.6 | M149[d] | 82 | 94.6 |
| T316 | 168 | 96.0 | M245 | 110 | 90.6 |
| T51 | 150 | 89.2 | J606 | 117 | 92.0 |

[a] The repetitive yields were calculated using yield data from gas chromatographic analysis at steps 2 and 21. [b] The yield in nanomoles at step 2 for each protein. [c] M63 was run twice. [d] Proteins analyzed at NIH.

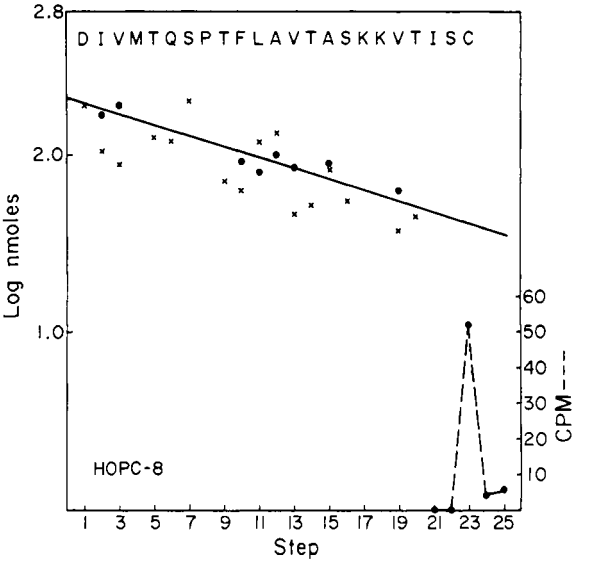


FIGURE 1: Yield data for H8. Yields are calculated by correction against the internal standard, norleucine (see text). (●) Yield of PTH derivatives analyzed on the gas chromatograph. (×) Yield of free amino acids after hydrolysis of the PTH derivatives and analysis on the amino acid analyzer. The one letter abbreviation of the amino acids (Dayhoff, 1969) is given above the respective yield points. D, aspartic acid; I, isoleucine; V, valine; M, methionine; T, threonine; Q, glutamine; S, serine; P, proline; F, phenalanine; L, leucine; A, alanine; K, lysine; and C, $^1/_2$-cystine. Dotted line indicates the counts in steps 21–25 due to [$^{14}$C]carboxamidomethylcysteine. A background of about 5 cpm is subtracted.

Lys from Arg, Thr from Pro, and also separates Ser, Gly, and Thr.

In some cases the remainder of the sample ($^{20}/_{25}$ths) was submitted to acid hydrolysis in 5.7 N HCl at 150° for 20 hr (Van Orden and Carpenter, 1964) or in 65% HI in an autoclave at 25 psi at 125° for 18 hr (Smithies *et al.*, 1971).[3] Free amino acids were generally analyzed on a Beckman 120 amino acid analyzer. Certain of the HCl hydrolysates were examined qualitatively by high-voltage electrophoresis on paper (Dreyer and Bynum, 1967).

Half-cystine residues were unambiguously identified by prior labeling with [$^{14}$C]iodoacetamide and scintillation counting of appropriate sequenator steps.

*PTH Sample Quantitation.* PTH-norleucine (250-nmol) was added as an internal standard to the tubes in which the thiazolinone derivatives were collected.[3] Individual sequenator steps were quantitated by adjusting the internal standard to 250 nmol and by correcting the other residues accordingly to compensate for sample loss throughout the many handling steps of this procedure. This procedure was used to quantitate the samples subjected to amino acid analysis as well as those submitted to gas chromatography.

*Peptide Map Analysis.* Light chains were alkylated with iodoacetamide or ethylenimine and digested with trypsin or thermolysin as previously described (Grant and Hood, 1971). Fingerprints were prepared according to the procedure of Katz *et al.* (1959).

*Nomenclature.* The tumors and their proteins are abbreviated in the text by the first letter of their name and their number, *e.g.*, MOPC 63 is designated M63. The complete tumor designations are given in Figure 2.

## Results

*Automatic Sequence Analysis.* The yield data for a typical protein, H8, are presented in Figure 1. The yield at each step is corrected, in part, for hydrolysis degradation by use of an internal standard, PTH-Nle, and is plotted as the log recovery in nanomoles of the predominant amino acid residue. Yields obtained from two methods of sample analyses, gas chromatography and amino acid analysis, were generally concordant (Figure 1). On occasion sharp breaks in the yield curve were noted between five and six. Presumably this represents sample loss due to partial cyclization of the Gln residue at position six to a pyrrolidonecarboxylic acid residue lacking a free α-amino group which is, accordingly, unavailable for sequenator analysis.

Also given in Figure 1 are the radioactive counts for samples in the vicinity of position 23. The $^1/_2$-Cys residue at position 23 is evident. The drop in radioactivity for position 24 indicates that less than 10% of the molecules are out of phase at this point in the sequenator run.

The yield data for this protein were typical of the others. Yield data for many of the remaining proteins were determined from gas chromatography analysis of the residues at positions 2, 11, and 21 (Leu or Ile).[3] Most of these curves were essentially linear when plotted as with Figure 1. The repetitive

[3] The sequenator methodology evolved markedly during the course of this study. In addition there was a change in instrumentation as one of us moved from NIH to Caltech (L. H.). The earlier proteins studied were not run with an internal standard; thus some of the recovery results are not quantifiable. The gas chromatograph used at NIH was a Varian Model 1840. At NIH all PTH-amino acids were hydrolyzed in 6 N HCl at 150° for 20 hr (Van Orden and Carpenter, 1964) for quantitative or qualitative amino acid analysis. This hydrolysis procedure destroys PTH-Ser, PTH-Thr, PTH-Trp, and PTH-Cys. Those proteins analyzed at NIH are indicated in Figure 2.

| Tumor No. | Class | 1 | 5 | 10 | 15 | 20 | 23 | Ref |
|---|---|---|---|---|---|---|---|---|
| MOPC 70 | IgF | Asp Ile Val Leu Thr Gln Ser Pro Ala Ser Leu Ala Val Ser Leu Gly Gln Arg Ala Thr Ile Ser Cys | | | | | | 1 |
| MOPC 321 | BJ | ——— | | | | | * | |
| TEPC 124 | BJ | ——— | | | | | * | |
| MOPC 63 | BJ | Asn——— | | | | | | |
| BFPC 32 | BJ | ——— | | | | | | |
| BFPC 61 | BJ | ——— | | | | Glx Lys——— | | |
| MOPC 46 | BJ | Asp Ile Val Leu Thr Gln Ser Pro Ala Thr Leu Ser Val Thr Pro Gly Asp Ser Val Ser Leu Ser Cys | | | | | | 2 |
| MOPC 172 | IgH | ——— | | | | | | 2 |
| MOPC 316 | BJ | ———( )———( )———(Thr)—Asx( )—( )Ile——— | | | | | | |
| MOPC 167+ | IgA | Asp Ile Val Ile Thr Gln Asx Glu Leu Ser Asp Pro Val Thr Ser Gly Glu Ser Val Ser Ile Thr Cys | | | | | | |
| MOPC 511+ | IgA | ——— | | | | | | 2 |
| McPC 603+ | IgA | Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ser Val Ser Ala Gly Glu( ) Val Thr Met Ser Cys | | | | | | |
| McPC 870+ | IgA | ———Lys——— | | | | | | |
| MOPC 384+ | IgA | ———Lys——— | | | | | | |
| LPC 1 | IgG | Asp Ile Val Met Thr Gln Ser Pro Ser Ser Met Gln Ala Ser Ile Gly Glu Lys Val Thr Ile Ser Cys | | | | | | |
| TEPC 15− | IgA | Asp Ile Val Met Thr Gln Ser Pro Thr Phe Leu Ala Val Thr Ala Ser Lys Lys Val Thr Ile Ser Cys* | | | | | | |
| HOPC 8− | IgA | ——— | | | | | * | |
| MOPC 35 | BJ | Asp Ile Val Met Thr Gln Thr Pro(Asx)Phe Leu Leu Val Ser Ala Gly( ) Arg Val Thr Ile Thr Cys | | | | | | |
| TEPC 157 | BJ | Asp Ile Val Met Thr Gln Ser Gln Ser Phe Met Ser Thr Ser Val Gly Asp Arg Val Ser Val Thr Cys | | | | | | 2 |
| MOPC 379 | BJ | Asp Ile Val Met Thr Gln Ser Pro Ser Phe Met Val(Thr)Ser Val Gly(Glx)( ) Val | | | | | | |
| HOPC 5 | IgA | Asp Ile Val Met Thr Gln Ser Thr Lys Phe Met Ser Thr Ser Val Gly(Lys) Arg Val Ser Ile Thr Cys | | | | | | |
| MOPC 21 | IgF | Asn Ile Val Met Thr Gln Ser Pro Lys Ser Met Ser Met Ser Val Gly Glu Arg Val Thr Leu Thr Cys | | | | | | 3 |
| MOPC 467+ | IgA | Asp Val Leu Met Thr Gln Thr Pro Leu Ser Leu Pro Val(Ser) Leu Gly Asp Glu Ala( ) Ile (Ser) Cys | | | | | | |
| MPC 37+ | IgH | ———( )———( )———(Ser)——— | | | | | | |
| TEPC 173 | BJ | Asp Ile Gln Met Thr Gln Ser Pro Ala Ser Leu Ser Val Ser Val Gly Glx Thr Val Thr Ile Thr Cys | | | | | | |
| RPC 23+ | IgF | ———Glu( )——— | | | | | | |
| MOPC 31c | IgF | Asp Ile Gln Met Thr Gln Ser Pro Ala Ser Leu Ser Ala Ser Val Gly Glu Arg Val Thr Ile Thr Cys | | | | | | |
| M | | ———Val———( )— | | | | | | |
| MOPC 173 | IgG | Asp Ile Gln Met Thr Gln Thr Thr Ser Ser Leu Ser Ala Ser Leu Gly Asp( ) Val Thr Ile Thr Cys | | | | | | |
| MOPC 41 | BJ | Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Leu Gly Glu Arg Val Ser Leu Thr Cys | | | | | | 1 |
| J606A | IgA | Asp Val Gln Met Ile Gln Ser Pro Ser Ser Leu Ser Ala Ser Leu Gly Asp Arg Val Thr | | | | | | |
| MOPC 149 | BJ | Asp Ile Gln Met Thr Gln Ser Pro Asx Tyr Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr( ) | | | | | | |
| McPC 600 | IgF | Asp Ile Gln Met Ile Glx Ser Pro Ser Ser Met Phe Ala Ser Ile Gly Asp Glx Val Ser Ile Ser Cys | | | | | | 2 |
| McPC 674 | BJ | Asp Val Val Met Thr Gln Thr Pro Leu Thr Leu Ser Val Thr Ile Gly Glu Pro Ala Ser Leu Ser Cys | | | | | | |
| McPC 843 | BJ | ———( )Ile——— | | | | | | |
| MOPC 460 | IgA | Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Thr Val Ser Leu Gly Asp Arg Ala Ser Ile Ser Cys* | | | | | | 4 |
| MOPC 265 | BJ | Glu Thr Thr Val Thr Gln Ser Pro Ala Ser Leu Ser Met Ala Ile Gly Glu Lys Val Thr Ile [+] Cys* | | | | | | |
| McPC 773 | BJ | ——— | | | | | * | |
| SAPC 10 | IgA | Glu Ile Val Leu Thr Glx Ser Pro Ala Ile Thr Ala Ala Ser Leu Gly Glx Arg Val Thr Ile Thr Cys | | | | | | 5 |
| TEPC 191 | | ——— | | | | | | 5 |
| MOPC 29 | BJ | Glu Asn Val Leu Thr Glx Ser Pro Ala Ile Met Ser Ala Ser Pro Gly Glu Arg Val Thr Met Thr Cys | | | | | | 2 |
| TEPC 153 | BJ | ———( )———Glx Asx——— | | | | | | |
| TEPC 29 | BJ | Glu Val Val Leu Thr Glx Ser Pro Ala Ile Met Ser Ala Ser Leu Gly Leu Arg Val Ser Met Ser Cys | | | | | | |
| MPC 47− | IgF | Glu Val Val Met Thr Gln Thr Pro Leu Ser Leu Ala Val (Ser) Leu Gly ( ) Glx Ala(Ser) | | | | | | |

FIGURE 2: Amino-terminal sequences of κ chains from inbred BALB/c mice. Tumor number designates the myeloma tumor from which the respective light chain was derived. Class indicates the immunoglobulin class (Potter, 1967). BJ designates a Bence-Jones protein. (*) The ½-cystine residue was confirmed by alkylation with [$^{14}$C]iodoacetamide. (+) Proteins only analyzed at NIH. (++) Tumor number of this protein was lost, but the sequence is included as it is different from each of the other proteins. Parentheses with or without residues indicate an uncertainty in residue assignment (see text). Brackets indicate that a gap must be introduced to maintain the homology position of cysteine at position 23. References: 1, Gray *et al.*, 1967; 2, E. Appella, personal communication; 3, Milstein and Svasti, 1971; 4, Haimovich, Eisen, and Hood, unpublished observation; and 5, S. Rudikoff, personal communication.

yields at each step calculated from the yields at position 2 and 21 are presented in Table II.

The amino-terminal sequences for 44 BALB/c κ chains are given in Figure 2. Those proteins which differed by two residues or less are listed together. The amino-terminal sequences of six κ chains from BALB/c hybrids are presented in Figure 3. Two comments should be made regarding these sequences. First, the ½-Cys residues for a limited number of proteins were determined using the radioactive labeling technique and these are indicated in Figure 2. Since the remainder of the proteins had residues at position 23 which were consistent with ½-Cys by gas chromatography and amino acid analysis, a tentative assignment was made. This tentative assignment is supported by the extensive degree of sequence conservation seen in this region for the κ chains of all other animals examined (Smith *et al.*, 1971) as well as by the homology seen in the more completely sequenced κ mouse chains (Gray *et al.*, 1967; Milstein and Svasti, 1971).[4] Second, the justification for the publication of proteins with uncertain residue assignments (blank positions) should be made explicit. They are included because these data taken as a whole have interesting genetic implications (see Discussion) which are independent of the uncertain residue positions. The proteins with blanks were generally available in quantities sufficient for but a single sequenator run. Often the yields for these proteins were low, making it difficult to detect the less stable PTH derivatives (*e.g.*, Ser and Thr). These low yields may be due to the

[4] The dangers in making such tentative assignments are evident in proteins M265 and M773. Initially a deletion at position 22 was missed (Hood *et al.*, 1970) because [$^{14}$C]iodoacetamide was not used to verify the tentative ½-Cys residue at position 23.

| Tumor No. | Class | Hybrid Partner | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MOPC 320 | A | C57BL | Asp | Ile | Gln | Met | Thr | Gln | Ser | Pro | Ser | Ser | Leu | Ser | Ala | Ser | Leu | Gly | Glx | Arg | Val | Ser | Leu | Thr | Cys |
| GPC 7 | H | NZB | Asp | Ile | Val | Met | Ser | Gln | Ser | Pro | Ser | Ser | Leu | Ala | Val | Ser | Ala | Gly | Glu | ( | ) | Val | Thr | Met | Ser | Cys |
| MOPC 512 | F | C57BL | Asp | Ile | Val | Leu | Thr | Gln | Thr | Pro | Ala | Ser | Ile | Thr | Ala | Ser | Leu | Lue | Gly | Glx | Val | Val | Thr | | | |
| MOPC 245 | F | C57BL | Asp(Ile) | Val | Met | Thr | Gln | Ser | Pro | (Ser) | Phe | Met | Ser | Thr | Ser | Val | Gly | Asn(Arg) | Val | Ser | Val | Thr | Cys | | | |
| MOPC 461 | IgH | C57BL | Asp | Ile | Gln | Met | Thr | Gln | Thr | Thr | Ser | Ser | Leu | Ser | Ala | Ser | Leu | Gly | ( | ) | Val | Thr | Ile | Thr | Cys | |
| MOPC 300 | F | C57BL | Asp | Val | Val | Leu | Thr | Gln | Thr | Pro | Leu | Ser | Leu | Thr | Ala | Ser | Leu | | | | | | | | | |

FIGURE 3: Amino-terminal sequences of κ chains from BALB/c hybrids. The GPC7 protein is from an F1 mouse of an (NZB × BALB/c) cross. The other hybrids have been produced by introgressively crossing the C57BL heavy chain locus onto a BALB/c background (see Potter, 1967, for details). M320 has been back-crossed six times; M512, four times; M300, six times; M245, three times; and M461, six times. GPC7 was a gift from L. Herzenberg. See legend to Figure 2 for details.

fact that most of these light chains were isolated in urea solutions and partial loss of free α-amino groups may have resulted through its reaction with cyanate ion. In addition, the sequences of most of the "blank" proteins were determined at NIH before analytic methods had been developed.[3]

A final comment should be made about sequence artifacts due to deamidation. The amino-terminal sequences of M321, T124, M70, and M63 are identical apart from a single amide difference. The amino-terminal residue of M63 was *ca.* 75% Asn and *ca.* 25% Asp as determined qualitatively by thin layer chromatography. The other three proteins had only Asp at their amino terminus by thin-layer chromatography. It is possible that deamidation of mouse light chains can occur postsynthetically in the mouse or during light-chain purification. A similar partial deamidation has been reported for M21 (Milstein and Svasti, 1971).

*Peptide Map Analysis of Light Chains with Identical Amino-Terminal Sequences.* The sets of proteins with identical amino-terminal sequences were cleaved with thermolysin and trypsin after reduction and alkylation and these digests were subjected to comparative fingerprint analysis. These proteins fell into two groups. First, H8 and T15 appeared to be identical by both procedures, as were M773 and M265 (Figure 4 and Table III). Thus these proteins are very similar, if not identical. Second, light chains M321, M63, M70, and M124 had minor peptide differences as did proteins M603, M870, and M384 and proteins M674 and M843 (Table III). The peptide map comparisons for one set of proteins with limited differences are given in Figure 5. Thus these proteins were similar, but had distinct sequence differences.

TABLE III: Peptide Map Comparisons of Proteins with Identical Amino-Terminal Sequences.[a]

| Proteins Compared | Enzymes: Trypsin[b] | Enzymes: Thermolysin | Indistinguishable[c] | Similar[c] |
|---|---|---|---|---|
| M773-M265 | + | + | + | |
| H8-T15 | + | + | + | |
| M70-M321-M124-M63 | + | | | + |
| M603-M870-M384 | + | | | + |
| M674-M843 | + | | | + |

[a] M674-M843 differ by one residue and M63 apparently differs from the other proteins in its group by an amide group. [b] All proteins were aminoethylated before trypsin digestion. [c] Indistinguishable proteins had no detectable peptide differences. Similar proteins differed by two to four peptides.

## Discussion

κ chains from the inbred BALB/c mice were studied to eliminate genetic polymorphisms which might obscure sequence patterns produced by the mechanism which generates antibody diversity. This diversity has been compared against that of the more throughly studied human κ family. Two generalizations emerge from this comparison. First, mouse κ chains exhibit more sequence diversity than their human counterparts, despite the fact that the mice used were highly inbred as compared to humans. Second, the mouse sequences are quite distinct from their human counterparts.

*BALB/c κ Chains from Myeloma Tumors Exhibit more Sequence Diversity Than Their Human Counterparts.* Forty-five

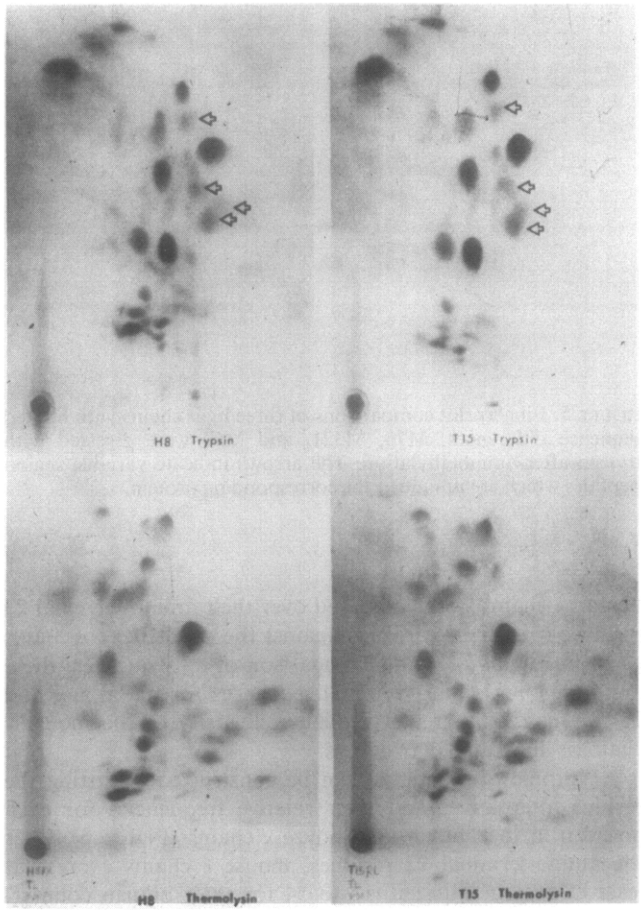


FIGURE 4: Fingerprint comparisons of two light chains with possible sequence identity. After total reduction and alkylation, the proteins H8 and T15 are digested with trypsin (upper panel) or thermolysin (lower panel). Four variable region peptides with identical mobilities are indicated by arrows in the trypsin comparison.

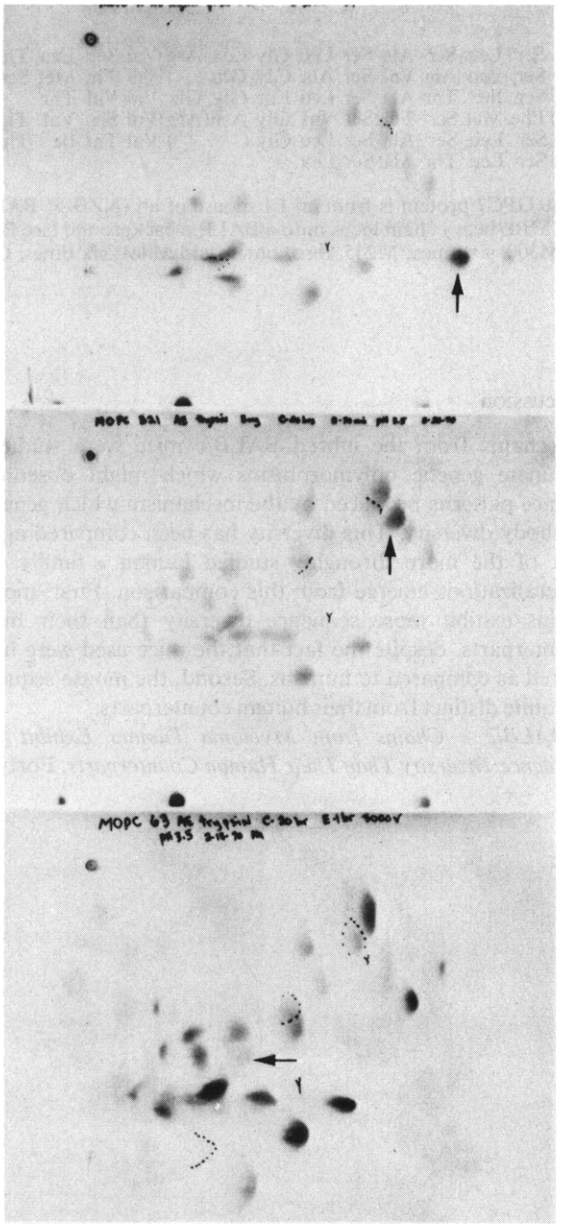

FIGURE 5: Fingerprint comparisons of three light chains with limited sequence differences. M70, M321, and M63 were digested with trypsin after aminoethylation. The arrows indicate variable region peptides which are unique to the corresponding protein.

human κ chains were compared over their amino-terminal 23 residues (Smith *et al.*, 1971) against the 44 BALB/c κ chains given in Figure 2. In this comparison of 89 sequences, there are 32 distinct amino-terminal sequences for the mouse and 26 for man. The greater sequence diversity of the mouse κ chains is illustrated in two ways.

A "profile of diversity" can be constructed by listing the residue alternatives and their relative frequencies for each position in the human and mouse κ chains (Figure 6). Over the amino-terminal 22 residues, mouse κ chains express at least 88 residue alternatives to 54 for their human counterparts. Fourteen positions in mouse κ chains show more diversity than human chains, whereas only four have less. In mouse κ chains, five positions have six or more alternatives (*e.g.*, 9, 10, 12, 15, 18), whereas only one position in human κ chains exhibits more than four residues (Figure 6). Thus mouse κ chains are significantly more diverse than their human counterparts at the amino terminus of these chains.

More important, most of the distinct sequences for the N terminus of the mouse κ chains ($^{32}/_{44}$ths) differ from one another far more than do their human counterparts differ from one another. This can be illustrated by constructing a genealogic tree for the amino-terminal 22 residues (Fitch and Margoliash, 1967). In Figure 7 the first 15 $V_\kappa$ sequences analyzed from the mouse are used as well as 35 human sequences. This genealogic analysis depicts the minimum number of genetic events (*e.g.*, gene duplication, sequence insertions or deletions, and single base substitutions) required to generate a given set of sequences (*e.g.*, $V_\kappa$ regions of human and mouse) from a single ancestral gene. This genealogic tree illustrates that the human κ chains fall onto three branches or "subgroups," $V_{\kappa I}$, $V_{\kappa II}$, and $V_{\kappa III}$. The human sequences within a subgroups generally differ by O-3 residues. In contrast, the 15 mouse $V_\kappa$ sequences fall onto a least nine major branches, $V_{\kappa I}$ to $V_{\kappa IX}$, which differ by 7–16 nucleotide substitutions. Admittedly, this definition of a major branch or subgroup is arbitrary (*e.g.*, 7 or more nucleotide substitutions in the N-terminal 22 residues). Note, however, that this tree is only based on residues constituting about one-fifth of the $V_\kappa$ region. If we assume that the diversity at the amino terminus is representative on the entire V region (which is true of human $V_\kappa$ regions), seven nucleotide substitutions at the amino terminus would indicate that there are about 35 substitutions throughout the entire V region. This is the lower limit of the differences noted in comparing two human proteins of differing subgroups and accordingly 7 nucleotide substitutions in 22 residues seems a reasonable limit for the difference required of major genealogic branches (Figure 7). Examination of the sequence data for the next 16 BALB/c proteins studied shows that they are, if anything, more diverse than the first set of 15 (Figure 2). Thus there are many more major branches (*e.g.*, subgroups) in the mouse κ genealogic tree than in its human counterpart.[5]

Another approach to the definition of major $V_\kappa$ region branches is the use of sequence gaps. Unfortunately this approach requires complete V region sequence data. It is worth stressing, however, that three sequence gaps are evident in the limited mouse κ chain data now available. M70 has an insertion of four residues when compared to M41 (Gray *et al.*, 1967). M265 and M773 have a deletion of one residue at position 22 when compared to all other $V_\kappa$ regions (Figure 2). It appears likely that additional sequence gaps will be defined as more complete $V_\kappa$ sequences become available.

*Are There More Mouse $V_\kappa$ Genes Than Human $V_\kappa$ Genes?* Obviously questions about the numbers of $V_\kappa$ genes can only be approached indirectly through the amino acid sequence analysis of antibody polypeptide chains. Two lines of evidence suggest that there may be more mouse $V_\kappa$ genes than human $V_\kappa$ genes. First, as noted above there are more distinct branches (*e.g.*, V region subgroups) on the mouse $V_\kappa$ genealogic tree than on its human counterpart. Most immunologists agree that each distinct subgroup represents at least one germ line V gene (Cohn, 1971; Gally and Edelman, 1970; Milstein and

[5] All attempts to place the mouse $V_\kappa$ sequences in a limited number of subgroups have failed. A matrix of minimum nucleotide differences for these proteins shows variation which ranges continuously between 0 and 16 nucleotides. An attempt on the computer to use the amino terminal four residues for subgroup classification (as can be done with human κ chains) also failed.

| Species | Position 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Human | D34 | I43 | Q27 | M31 | T45 | Q45 | S44 | P45 | S26 | S28 | L45 | S38 | A23 | S36 | V26 | G44 | D30 | R38 | V26 | T34 | I33 | T24 |
| | E10 | M 1 | V18 | L14 | | | T 1 | | G 8 | T16 | | P 7 | L10 | T 7 | P18 | R 1 | E15 | P 7 | A18 | S 8 | L12 | S18 |
| | K 1 | V 1 | | | | | | | L 7 | P 1 | | | V 9 | V 2 | L 1 | | | | I 1 | A 1 | | A 2 |
| | | | | | | | | | A 4 | | | | T 2 | | | | | | | | | Z 1 |
| | | | | | | | | | | | | | M 1 | | | | | | | | | |
| BALB/c | D34 | I32 | V31 | M26 | T42 | Q44 | S34 | P39 | A20 | S27 | L31 | S24 | V26 | S33 | L16 | G42 | E18 | R17 | V32 | T26 | T29 | S21 |
| Mouse | E 8 | V 8 | Q 9 | L14 | I 2 | | T 8 | T 2 | S10 | F 6 | M 9 | A11 | A12 | T 9 | V10 | S 2 | D 8 | K 8 | A12 | S14 | L 5 | T15 |
| | N 2 | T 2 | T 2 | I 2 | | | B 2 | E 2 | L 8 | T 5 | D 2 | P 4 | M 3 | A 2 | I 6 | | Z 6 | S 4 | | | M 5 | [ ]2 |
| | | N 2 | L 2 | V 2 | | | | Q 1 | T 2 | I 4 | T 2 | F 1 | T 3 | | A 6 | | Q 5 | T 2 | | | V 1 | |
| | | | | | | | | | K 2 | L 1 | | L 1 | | | P 3 | | K 3 | E 2 | | | | |
| | | | | | | | | | B 2 | Y 1 | | T 1 | | | S 2 | | L 1 | P 2 | | | | |
| | | | | | | | | | | | | V 1 | | | T 1 | | B 1 | Z 2 | | | | |
| | | | | | | | | | | | | Q 1 | | | | | | L 1 | | | | |
| | | | | | | | | | | | | | | | | | | B 1 | | | | |

FIGURE 6: A profile of sequence diversity for the myeloma κ chains of mouse and man. Forty-five human chains (Smith *et al.*, 1971) with complete amino-terminal sequences for 22 or more residues are compared against the 44 mouse κ chains in Figure 2. The residue alternatives at each position are listed along with their relative frequencies. Major residues (eight or more proteins) present in just one of these species are boxed. These frequencies are not converted to percentages as the number of proteins compared in both systems is similar. Amide or acid assignments were made by homology where possible. The one letter amino acid code is used: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr; and Z, Glx.

Pink, 1970; Smith *et al.*, 1971; Hilschman *et al.*, 1970). Second, the ratio of κ:λ chains is 97:3 in the serum of the mouse (McIntire and Rouse, 1970) and 2:1 in man (Fahey, 1963). Cohn and his coworkers have suggested that the serum ratios of immunoglobulin chains may indirectly reflect the number of germ line V genes in their respective immunoglobulin families (Cohn, 1971). This supposition is consistent with the observation that mouse $V_\lambda$ regions, present in the serum at the 3% level, exhibit very limited sequence diversity (Weigert *et al.*, 1970). Perhaps the mouse may have had to expand its $V_\kappa$ gene population to compensate for a contraction in the number of $V_\lambda$ genes.

The argument developed in the preceding paragraph may have to be qualified in that the myeloma process in man and mouse does not appear to select random immunocytes for transformation (Potter, 1973). First, tumor induction occurs at different sites (peritoneal cavity for mouse and probably the bone marrow for man). Second, IgA is the predominant class of myeloma proteins in the mouse whereas IgG is the predominant class in man. Finally, mouse myeloma proteins with antibody-like activity appear to exhibit a very limited spectrum of activities (Potter, 1973). Thus caution must be used in translating phenotype (myeloma proteins or serum immunoglobulin ratios) to genotype (number of V genes). Nevertheless, it does appear that the myeloma system can be used to estimate the *lower limits* of antibody diversity. If the nature of myeloma selection is different in various species, the estimates of lower limits on V gene pools cannot be directly compared in different animals.

*The Diversity in BALB/c $V_\kappa$ Regions Appears to Be Encoded by Many Germ Line Genes.* One can see from the genealogic tree that the distinct branches on the mouse $V_\kappa$ tree differ by an average of nine or more nucleotide substitutions (Figure 7). Again, assuming that the diversity at the amino terminus is representative, an average of 45 substitutions will, on the average, separate the $V_\kappa$ branches for the entire variable region. The extent of this divergence suggests that each $V_\kappa$ branch should be encoded by a separate germ line gene (see Hood and Prahl, 1971). By this criterion more than 25 $V_\kappa$ genes should code for the sequences in Figure 2. In this regard it is worth emphasizing that more new branches (possible germ line V genes) were added to the genealogic tree for the last 16 mouse κ chains analyzed than were defined by the first 15 studied. Thus there is no evidence that we are approaching any limit on the diversity (branching) of the $V_\kappa$ genealogic tree of the BALB/c mouse. Accordingly, the minimum number of germ line $V_\kappa$ genes in the BALB/c mouse appears to be much greater than 25.

The pattern of limited sequence variation in mouse λ chains has led Weigert *et al.* (1970) to propose that somatic diversification occurs only in three very limited portions of the variable region (these areas are termed "hypervariable" regions). They suggest that each amino acid substitution which occurs outside these hypervariable regions is encoded by a distinct germ line gene. Since the amino-terminal 22 residues lie outside these hypervariable regions, each distinct amino terminal sequence should, by the Weigert hypothesis, be encoded by a separate germ line gene. The 44 sequences in Figure 2 would then represent 32 germ line genes, approximately 1 V gene/1.4 sequences studied. Although this possibility appears unlikely to us, the Weigert hypothesis also suggests that many $V_\kappa$ genes exist in the germ line of the BALB/c mouse.

*Mouse $V_\kappa$ Sequences Are Quite Distinct from Their Human Counterparts.* Figure 7 illustrates that mouse κ chains occupy branches on the genealogic tree of V regions which are distinct from those of their human counterparts (the mouse branches differ *ca.* 40–50% in sequence from their human counterparts

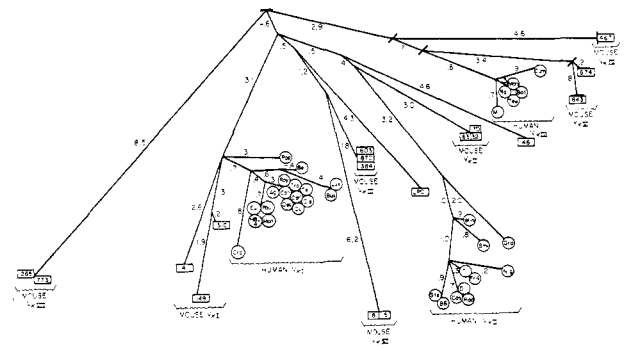

FIGURE 7: Genealogic tree for positions 1–22 of mouse and human κ chains. Constructed by the method of Fitch and Margoliash (1967) and adapted from Smith *et al.* (1971).

when complete V region comparisons are made).[6] This difference is also illustrated by the observation that major residue alternatives are associated with one species and not with a a second. For example, man has species associated residues in the $V_\kappa$ regions at positions 9 (Gly) and 13 (Leu) which are not found at these positions in the mouse, whereas the mouse has species associated residues at positions 11, 12, and 18 (boxed residues in Figure 6). The three species associated residues in the mouse generally are *not* linked together in the same sequences, implying that three sets of $\kappa$ genes have evolved in the mouse evolutionary line after its divergence from the human evolutionary line. The implications of these species associated differences with regard to the evolution of multiple $V_\kappa$ genes have been discussed elsewhere (Hood and Prahl, 1972).

At many positions the major residue alternatives are similar in human and mouse $\kappa$ chains (*e.g.*, aspartic acid-glutamatic acid at position 1; valine-glutamine at position 3, etc.). This conservation of sequence alternatives in both systems is consistent with the presence of great selective pressures on particular residue positions. However, these same highly selected residues are linked together in sequences which are characteristic for each species as illustrated by the distinctness of the mouse and human $V_\kappa$ branches on the genealogic tree (Figure 7). Thus individual residues are highly selected across species boundaries, but the sequences into which they are linked are distinct for each species.

The $V_\kappa$ sequences from myeloma light chains derived from six BALB/c hybrids are given in Figure 3. There are two kinds of hybrids. The GPC7 $\kappa$ chain comes from the F1 generation of an (NZB $\times$ BALB/c) cross. This light chain has a 50% chance of being derived from the BALB/c genome. The other hybrids were derived from mice in which a C57BL heavy-chain genome had been introgressively back-crossed into BALB/c genome (see Figure 3 for details). Thus, the light chains are probably of the BALB/c type.[7] Four of these sequences differ from the 44 $V_\kappa$ sequences in Figure 2 whereas MOPC 461 is identical with MOPC 245 and MOPC 329 is identical with MOPC 41. There are no features of the amino terminus which distinguish these chains from those of the inbred BALB/c mouse apart from Ser $\leftrightarrow$ Thr interchange at position 5 in the GPC7 chain which has not been previously reported. If these sequences are coded for by BALB/c $V_\kappa$ genes, 36 out of 50 of the sequences studied are unique.

*Myeloma Proteins with Shared Antibody Specificity Can Employ Distinct $V_\kappa$ Sequences.* Three groups of immunoglobulins with antibody activity have been studied. Proteins H8, T15, M167, and M603 have antibody activity to antigens that contain phosphorylcholine (Potter, 1970; Potter *et al.*, 1970). Proteins M870, M384, and M467 have activity against various Salmonella O antigens (Potter, 1970). Protein M460 has antibody activity against $N_2$ph (Eisen *et al.*, 1970). The light chains from these eight proteins fall into five distinct branches of the genealogical tree. Light chains from immunoglobulins that have at least two different antibody specificities are present in the M603, M870, M384 branch. Thus, no simple correlation between antibody activity and the amino-terminal sequence of the V region is apparent (see Kaplan and Metzger, 1969; Cohen and Copper, 1968). This observation underscores the degeneracy of the immune response; that is, a single antigen generally stimulates the production of many different molecular species of antibody molecules. Furthermore, antibody polypeptide chains of a single specificity can come from distinct branches (subgroups) on the genealogic tree.

Nine branches on the complete genealogical tree have two or more light chains whose amino-terminal sequences may be identical (see identical sequences in Figure 2). These proteins fall into two groups. First, light chains in branches containing T15-H8 and M265-M773 appear to be very similar if not identical by a variety of criteria–thermolytic and tryptic fingerprints (Figure 4) and immunologic analysis with reagents exhibiting individual specificity (Potter and Leiberman, 1970; P. Periman, unpublished results). These proteins will be the subject of a future report. The second group includes proteins with identical amino-terminal sequences and minor but unequivocal V region peptide differences on tryptic fingerprints (Table III and Figure 5). A study of proteins with limited sequence difference is the subject of two companion reports (McKean *et al.*, 1972a,b).

Acknowledgment

We thank Drs. E. Appella and S. Rudikoff for permission to use his unpublished sequence data.

References

Cohen, S., and Cooper, A. (1968), *Immunology 15*, 93.
Cohen, S., and Milstein, C. (1967), *Advan. Immunol. 7*, 1.
Cohn, M. (1971), *Ann. N. Y. Acad. Sci. 190*, 529.
Dayhoff, M. (1969), Atlas of Protein Sequence and Structure, Silver Spring, Md., National Biomedical Research Foundation.
Dreyer, W. J., and Bynum, E. (1967), *Methods Enzymol. 11*, 32.
Edelman, G. M., and Gall, W. E. (1969), *Annu. Rev. Biochem. 38*, 415.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.
Eisen, H., Michaelides, M. C., Underdown, B. J., Schulenberg, E. P., and Simms, E. S. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 78.
Fahey, J. L. (1963), *J. Immunol. 91*, 483.
Fitch, W. F., and Margoliash, E. (1967), *Science 155*, 279.
Gally, J., and Edelman, G. M. (1970), *Nature (London) 227*, 341.
Grant, J. A., and Hood, L. (1971), *Biochemistry 10*, 3123.
Gray, W. R., Dreyer, W. J., and Hood, L. (1967), *Science 155*, 465.
Hilschmann, N., Ponstingl, H., Baczko, K., Braun, D., Hess, M., Suter, L., Barnikol, H. U. and Watanabe, S. (1970), *Protides Biol. Fluids Proc. Colloq. 17*, 53.
Hood, L. (1972), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 31*, 177.
Hood, L., McKean, D., and Potter, M. (1970), *Science 170*, 1207.
Hood, L., and Prahl, J. (1971), *Advan. Immunol. 14*, 291.
Hood, L., and Talmage, D. W. (1970), *Science 168*, 375.
Kaplan, A., and Metzger, H. (1969), *Biochemistry 8*, 3944.
Katz, A. M., Dreyer, W. J., and Anfinsen, C. B. (1959), *J. Biol. Chem. 234*, 2897.
McIntire, K. R., and Rouse, A. M. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 704.

[6] The apparent similarity of the N terminus of mouse proteins M149, M41, and M316 and human $V_{KI}$ proteins is deceptive as when the entire V region sequences are compared, these two branches are about as distinct as those of other mouse and human light chains (Hood and Talmage, 1970).

[7] This is true provided the $\kappa$ and H genes of the mouse are unlinked. The $\kappa$ and H genes of rabbit and man are unlinked. It is assumed that those of mouse will be also.

McKean, D., Potter, M., and Hood, L. (1973a), *Biochemistry 12*, 749.
McKean, D., Potter, M., and Hood, L. (1973b), *Biochemistry 12*, 760.
Milstein, C., and Pink, J. R. L. (1970), *Progr. Biophys. Mol. Biol. 21*, 209.
Milstein, C., and Svasti, J. (1971), *Progr. Immunol. 1*, 33.
Pisano, J., and Bronzert, T. (1969), *J. Biol. Chem. 244*, 5597.
Potter, M. (1967), *Methods Cancer Res. 2*, 105.
Potter, M. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 85.
Potter, M. (1973), *Proc. 3rd Int. Convocation Immunol.* (in press).
Potter, M., and Kuff, E. L. (1964), *J. Mol. Biol. 9*, 537.
Potter, M., and Lieberman, R. (1970), *J. Exp. Med. 132*, 737.
Potter, M., Lieberman, R., Hood, L., and McKean, D. J. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 437.
Smith, G. P., Hood, L., and Fitch, W. M. (1971), *Annu. Rev. Biochem. 40*, 969.
Smithies, O., Gibson, D. M., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10*, 4912.
Van Orden, H. O., and Carpenter, F. H. (1964), *Biochem. Biophys. Res. Commun. 14*, 399.
Weigert, M. G., Cesari, I. M., Yonkovich, S. J., and Cohn, M. (1970), *Nature (London) 228*, 1045.
World Health Organization (1969), *Bull. W. H. O. 41*, 975.

# Mouse Immunoglobulin Chains. Partial Amino Acid Sequence of a κ Chain†

David McKean,‡ Michael Potter, and Leroy Hood*,§

ABSTRACT: The nearly complete amino acid sequence from the variable region (residues 1–107) of BALB/c mouse κ chain MOPC 321 is presented. Compositional data are presented on tryptic, thermolytic, and chymotryptic peptides derived from the constant region (residues 108–214). Methionine residues at positions 33 and 175 permitted this protein to be cleaved with cyanogen bromide into three fragments which were separated in part by gel filtration. The cyanogen bromide fragments were subsequently subjected to enzymatic cleavage and the resulting peptides were isolated on paper by chromatography and electrophoresis. The automatic sequenator was used to confirm the sequence of the amino-terminal 37 residues. The combination of fortuitous methionine placement, isolation of peptides from paper, and the sequenator analysis permitted this protein to be analyzed very rapidly.

The previous paper in this series demonstrated that certain mouse immunoglobulin κ chains[1] are identical or nearly identical in amino-terminal sequence for 23 residues (Hood *et al.*, 1973). A comparative peptide map analysis of one group of these mouse κ chains, M321, M63, and T124,[2] showed that each of these light chains is similar but not identical with the others. We report here the nature of amino acid sequence differences which exist within this group of similar mouse immunoglobulin κ chains. This paper is a report on the methodology employed to sequence these proteins and describes in detail the analysis of one protein, M321. The following paper (McKean *et al.*, 1973) will present the two additional partial sequences from this group of proteins and will discuss the implications of these sequence comparisons with regard to genetic and evolutionary mechanisms of antibody diversity.

## Materials and Methods

*Protein Isolation.* M321 was isolated from the urine of female BALB/c mice bearing the corresponding plasma cell tumor. The urine was clarified by centrifugation and exhaustively dialyzed against 0.1 M ammonium bicarbonate. The mouse Bence-Jones protein was separated from mouse urinary protein (MUP)[3] by ion-exchange chromatography on a DEAE-A25 column equilibrated with 0.05 M Tris-acetate buffer at pH 5.5 (Potter, 1967). The light chain is eluted with this buffer whereas the MUP is bound.

*Protein Purity.* The light-chain purity was determined by agar gel electrophoresis at pH 8.2 (Potter and Kuff, 1964) (see Figure 1) and by Ouchterlony analysis with rabbit anti-MUP antiserum.

*Protein Preparation.* Purified light chains were dissolved in 7 M guanidine-HCl–0.2 M Tris-HCl (pH 8.6) at a concentration of 20 mg/ml. Intrachain disulfide bonds were reduced with 50 mM dithiothreitol at 37° for 90 min. An equal volume of 0.11 M iodoacetamide dissolved in 7 M guanidine-HCl–0.2

† From the Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218 (D. M.), from the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014 (M. P.), and from the Division of Biology, California Institute of Technology, Pasadena, California (L. H.). *Received August 18, 1972.*

‡ Present address: Department of Medical Genetics, University of Wisconsin, Madison, Wis. 53706. This work was in part carried out at the National Cancer Institute while the senior author was a guest worker, in fulfillment for a doctoral dissertation.

§ Work supported in part by grants GB 27605 from the National Science Foundation and GM 06965 from the National Institutes of Health.

[1] The abbreviations used for immunoglobulin chains are those proposed by the World Health Organization: see *Bull. W. H. O. 41*, 975 (1969).

[2] Mouse immunoglobulin chains MOPC 321, MOPC 63, TEPC 124, and MOPC 70 are abbreviated M321, M63, T124, and M70, respectively.

[3] Abbreviations used are: MUP, mouse urinary protein; PTH, phenylthiohydantoin; dansyl, 1 dimethylaminonaphthalene-5-sulfonyl.