Biochem. J. (1978) **171**, 337–347
Printed in Great Britain

Sequence of the Full-Length Immunoglobulin κ-Chain of Mouse Myeloma MPC 11

By GEORGE P. SMITH
**Division of Biological Sciences, University of Missouri, Tucker Hall, Columbia, MO 65201, U.S.A., and Departments of Genetics and Medical Genetics, University of Wisconsin, Madison, WI 53706, U.S.A.*

(*Received* 12 *July* 1977)

MPC 11 mouse myeloma cells synthesize two immunoglobulin κ light chains, coded by two separate genes. One of these κ-chains has no variable region and is degraded intracellularly. The other is a full-length κ-chain containing both variable and constant regions; this chain is secreted, both by itself and combined with heavy chains in molecules of immunoglobulin G. This paper reports the amino acid sequence of the myeloma MPC 11 full-length κ-chain. The chain is unusual in having 12 extra residues at its *N*-terminus when its sequence is aligned with those of other mouse κ-chains; no other anomalies were found in its sequence.

The cells of the mouse myeloma tumour line MPC 11 are unusual in that they synthesize two immunoglobulin light chains, coded by two different genes (Kuehl & Scharff, 1974; Kuehl *et al.*, 1975; Rose *et al.*, 1977). One of these chains is a full-length κ-type light chain which is secreted from the cell, both by itself and combined with heavy chains to form molecules of immunoglobulin G. When the sequence of this chain is aligned with those of other mouse κ-chains, it is seen to carry 12 extra *N*-terminal residues (Smith, 1973*a*). This 12-residue segment does not seem to be an example of the *N*-terminal segments (which I will call 'leaders' in this paper) found on light chains when they are produced by translation of mRNA *in vitro* but absent from the chains when they are secreted by intact cells (Milstein *et al.*, 1972; Burstein & Schechter, 1977). In fact, the form of this myeloma MPC 11 κ-chain that is produced by translation of mRNA *in vitro* also has a normal leader preceding its abnormal 12-residue extra segment (Rose *et al.*, 1977). The second myeloma MPC 11 κ-chain, which is degraded intracellularly without detectable secretion, is aberrant in having no variable region. The form of this chain produced by translation of mRNA *in vitro* consists of the first 17 residues of a normal κ-chain leader joined directly to a normal κ-chain constant region (Rose *et al.*, 1977).

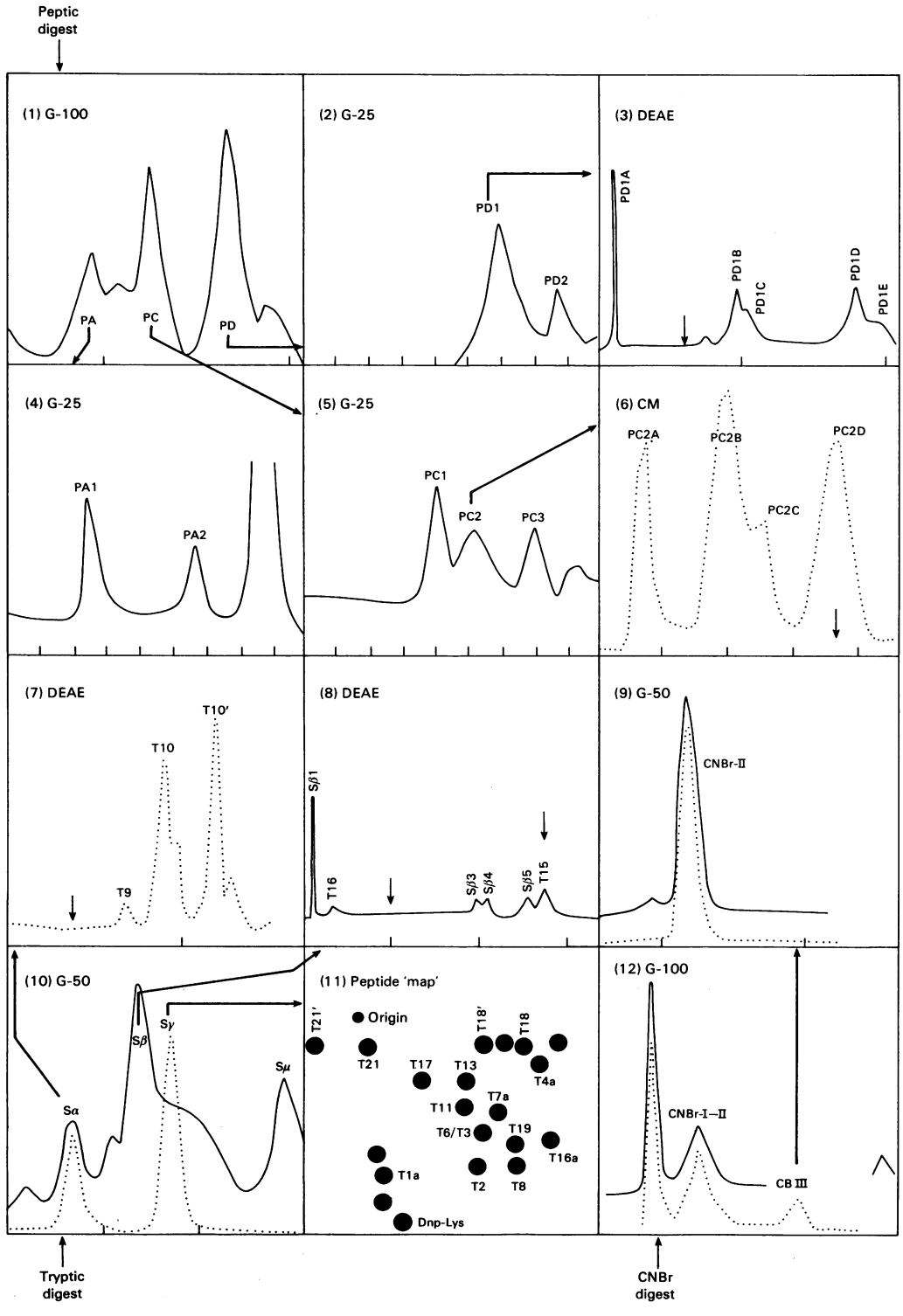
To ascertain if there are any other peculiarities of the myeloma MPC 11 κ-chain system, I have investigated the full-length κ-chains secreted from these cells more thoroughly. In this paper I report the amino acid sequence of the variable region and of most of the constant region, along with peptide-composition data covering all but four residues of the remainder of the chain. These studies have revealed no further abnormalities of this polypeptide.

* Present address.

Materials, Methods and Results

Isolation of MPC 11 immunoglobulin G

Peritoneal fluid was collected from mice that had been injected intraperitoneally with cells of myeloma MPC 11 clone 45.6 (Laskov & Scharff, 1970), and was freed of its 10–15% content of erythrocytes by centrifugation at about 100*g* for about 5min at about 24°C (the fluid very seldom clots). To 37ml of the supernatant was added 37ml of buffered saline (0.15M-NaCl/0.01M-NaH_2PO_4, pH adjusted to 7.0 with NaOH). Then 33ml of 3.5M-$(NH_4)_2SO_4$ (adjusted to pH approx. 7 with NaOH) was added gradually with stirring at 0°C; the light precipitate formed was removed by centrifugation at approx. 10000*g* for approx. 10min at 4°C. A further 26ml of 3.5M-$(NH_4)_2SO_4$ was added to the supernatant with stirring at 0°C, and the copious precipitate was collected by centrifugation as above. The precipitate was reprecipitated twice by dissolving it in buffered saline to a volume of 35ml, adding 35ml of 3.5M-$(NH_4)_2SO_4$ with stirring at 0°C, and collecting the precipitate by centrifugation as above. The final precipitate was dissolved in approx. 30ml of 0.05M-NaCl/0.1M-Tris, pH adjusted to 8.3 with HCl, dialysed for about 10h against each of two 4-litre volumes of the same buffer, and applied to a column (94cm × 2.5cm) of DEAE-Sephadex A-50 equilibrated with the same buffer. After extensive washing with the same buffer, the column was eluted with an asymptotic gradient (see under 'Ion-exchange chromatography: asymptotic-gradient elution'), the final buffer being 0.15M-NaCl/0.1M-Tris, pH adjusted to 8.3 with HCl, and the mixing-chamber volume being 250ml. The major peak, which emerged about 320ml after the beginning of gradient elution, was used as purified MPC 11 immunoglobulin G. The final yield


Peptic digest
(1) G-100
PA
PC
PD
(2) G-25
PD1
PD2
(3) DEAE
PD1A
PD1B
PD1C
PD1D
PD1E
(4) G-25
PA1
PA2
(5) G-25
PC1
PC2
PC3
(6) CM
PC2A
PC2B
PC2C
PC2D
(7) DEAE
T10'
T10
T9
(8) DEAE
Sβ1
T16
Sβ3
Sβ4
Sβ5
T15
(9) G-50
CNBr-II
(10) G-50
Sγ
Sβ
Sμ
Sα
(11) Peptide 'map'
Origin
T21'
T21
T17
T13
T18'
T18
T4a
T7a
T11
T6/T3
T19
T16a
T1a
T2
T8
Dnp-Lys
(12) G-100
CNBr-I–II
CB III
Tryptic digest
CNBr digest

was 19mg/ml of cell-free peritoneal fluid, or 70mg per mouse.

Sulphitolysis of immunoglobulin G and separation of heavy and light chains

Purified MPC 11 immunoglobulin G was concentrated to about 60mg/ml (A_{280} of 82) by vacuum dialysis, and was dialysed overnight against 2 litres of 0.5M-NH_4Cl, pH adjusted to 8.9 with NH_3. To 10ml of this solution was added 0.5ml of 0.1M-$CuSO_4$ and then 2.0ml of freshly prepared 1.25M-Na_2SO_3 (Cole, 1967). Sulphitolysis was allowed to proceed overnight at room temperature (about 24°C); the protein was then dialysed for approx. 2h against each of two 2-litre volumes of 0.2M-NH_4Cl (pH adjusted to 8 with NH_3). Formic acid was added to a pH of about 3.5 (indicator paper) and 0.4g of urea was added/ml of solution to give a final urea concentration of 5M. This solution was fractionated on two tandem columns (90cm × 2.5cm) of Sephadex G-100 in 5M-urea/0.2M formic acid/0.1M-ammonium formate. When, as in this case, more than about 50mg of protein (applied to the columns in a volume of 10–15ml) is fractionated in this manner, the chains tend to aggregate reversibly, causing marked skewing of the heavy- and light-chain peaks towards their leading edges, and necessitating re-running of the

Fig. 1. *Fractionation of peptides and peptide mixtures from the myeloma MPC 11 full-length κ-chain*

Heavy arrows indicate the source of the material fractionated in each step. Peptic, tryptic and CNBr digests were prepared as described in the text. The marks on the abscissae indicate the following effluent volumes in units of column volumes: (1), 10, 20 and 30; (2), (4) and (5), 0.1–0.8 in steps of 0.1; (3), 10 and 20; (6), 1, 2, 3 and 4; (7), 10; (8), 10, 20 and 30; (9) 0.5 and 1.0; (10) 0.7–1.1 in steps of 0.1. Effluent volumes were not recorded for panel 12, but the first peak probably consists of aggregates excluded from the gel and the inverted 'V' at the right shows the elution position of Phenol Red. The curves show profiles of the following measurements (in relative units) on column effluents: (1), A_{280}; (2)–(5) and (8), T_{280}; (6) and (7), radioactivity (c.p.m.); (9), ····, radioactivity (c.p.m.), ——, A_{220}; (10) and (12), ····, radioactivity (c.p.m.), ——, A_{280}. All absorbances and transmittances were measured in a cuvette or flow-cell with a 1-cm light-path. In the peptide 'map' in (11), chromatography was downward, electrophoresis was towards the right and left with the electrodes positioned as indicated by the '+' and '–', and 'Dnp-Lys' stands for ε-dinitrophenyl-lysine. In (3) and (6)–(8), the downward-pointing arrows mark the beginning of asymptotic-gradient elution (see the text) in ion-exchange chromatography. Details of the fractionations are given in the table below; the chromatographic media referred to are: G-100, Sephadex G-100; G-50, Sephadex G-50 (fine grade); G-25, Sephadex G-25 (fine grade); A-25, DEAE-Sephadex A-25; C-25, CM-Sephadex C-25; 3MM, Whatman 3MM chromatography paper.

Panel	Type	Medium	Column volume (ml)	(Initial) buffer	Final buffer	Mixing chamber volume (ml)
(1)	Gel filtration	G-100	900	5M-Urea/0.2M-formic acid/0.1M-ammonium formate	—	—
(2)	Gel filtration*	G-25	400	1% (v/v) Acetic acid	—	—
(3)	Ion exchange	A-25	16	0.02M-Ammonium acetate, pH adjusted to 8.1 with NH_3	0.4M-Ammonium acetate, pH adjusted to 8.1 with NH_3	100
(4)	Gel filtration*	G-25	400	1% (v/v) Acetic acid	—	—
(5)	Gel filtration*	G-25	400	1% (v/v) Acetic acid	—	—
(6)	Ion exchange	C-25	22	0.02M-Ammonium acetate, pH adjusted to 8.1 with NH_3	0.1M-Ammonium acetate, pH adjusted to 8.1 with NH_3	125
(7)	Ion exchange	A-25	16	0.02M-Ammonium acetate, pH adjusted to 8.1 with NH_3	0.4M-Ammonium acetate, pH adjusted to 8.1 with NH_3	100
(8)	Ion exchange	A-25	16	0.02M-Ammonium acetate, pH adjusted to 8.1 with NH_3	0.4M-Ammonium acetate, pH adjusted to 8.1 with NH_3 at second arrow: 0.85M-Ammonium acetate, pH adjusted to 8.1 with NH_3	100
(9)	Gel filtration	G-50	295	0.1M-NH_3	—	—
(10)	Gel filtration	G-50	850	1% (w/v) NH_4HCO_3	—	—
(11)	Peptide 'map'	3MM	—	—	—	—
(12)	Gel filtration	G-100	450	9% (w/v) Formic acid	—	—

* After complete reduction and *S*-[^{14}C]carboxamidomethylation as described in text.

Table 1. *Amino acid compositions and yields of tryptic and peptic peptides and peptide mixtures of MPC 11 full-length κ-chain*
Relative yields of amino acids in 4 M-methanesulphonic acid hydrolysates are given in roman type, under which are shown in italic type the corresponding 'true' compositions assumed in deducing the sequence. The peptide yields are given as percentage molar yield per mol of starting light chain.

	Content (mol of residue/mol of peptide)																		Yield
Peptide	Cys*	Asx	Thr	Ser	Glx	Pro	Gly	Ala	Val	Met	Ile	Leu	Tyr	Phe	His	Lys	Arg	Trp†	(%)
Tla		1.1 *1*	1.1 *1*	0.3	1.2 *1*		0.3		0.8 *1*	0.9 *1*	0.8 *1*		0.2	1.1 *1*					1.4
T2		1.0 *1*	1.0 *1*	1.7 *2*			1.0 *1*		1.0 *1*	1.0 *1*				0.9 *1*			1.0 *1*		10.8
T3/T6	1.0 *1.5*	1.0 *1*	2.1 *2.5*	1.3 *1.5*		1.1 *1.5*	1.0 *1*		2.5 *2.5*		1.5 *1.5*		0.7 *1*			1.4 *1.5*	1.0 *1*		5.6/8.1
T4a				0.9 *1*	1.0 *1*	1.9 *2*	1.0 *1*										1.0 *1*		8.4
T7a			2.8 *3*	0.8 *1*		2.1 *2*	3.1 *3*							1.0 *1*		1.0 *1*			8.0
T8					1.0 *1*						1.0 *1*	1.0 *1*				1.0 *1*			24.2
T10	0.4 *1*	2.9 *3*	1.6 *2*	2.8 *5*	2.1 *2*	3.9 *4*	2 5 *2*	4 0 *4*	2.8 *3*		1.0 *1*	2.0 *2*	0.5 *1*	2.9 *3*		1.0 *1*			20.3
T11		2 0 *2*							1.0 *1*		1.0 *1*					1.0 *1*			31.5
Sμ																1.0 *1*		+ *1*	
T13		1.1 *1*	0.1	1.0 *1*	0.9 *1*		1.1 *1*		0.1		0.9 *1*			0.1			1.1 *1*		28.9
T15		3.9 *4*	0.9 *1*	1.1 *2*	2.0 *2*		1.1 *1*		1.0 *1*			1.1 *1*				1.0 *1*		+ *1*	30.6
T16		1.0 *1*	3.8 *4*	3.3 *4*						0.6 *1*		2.1 *2*	0.9 *1*			1.1 *1*			40.0
T16a			1.5 *2*									1.0 *1*				1.0 *1*			2.1
T17		1.0 *1*			2.0 *2*								0.8 *1*				1.0 *1*		15.0
T18	0.8 *1*	1.0 *1*	1.8 *2*	0 9 *1*	1.1 *1*			1.1 *1*					0.8 *1*		1.6 *2*	1.0 *1*			4.4
T19			1.7 *2*	1.6 *2*		1.0 *1*			0.9 *1*		0.9 *1*					1.0 *1*			14.1
T21	0 8 *1*	1.0 *1*			1.0 *1*														14.8
PC2B	1.0 *1*		2.2 *3*	0.5 *1*	2.1 *2*	1.8 *2*	2.9 *3*					1.0 *1*	2.8 *3*	1.0 *1*	0.8 *1*	1.0 *1*			12.6
PC2D	1.0 *1*		2.6 *3*	0.9 *1*	2.2 *2*	1.8 *2*	2.8 *3*	1.0 *1*	0.8 *1*			1.0 *1*	2.7 *3*	1.0 *1*	0.8 *1*	0.9 *1*			9.9
PD1A				0.9 *1*	3.0 *3*	2.0 *2*	1.0 *1*	0.2	0.2			1.1 *1*	0.8 *1*			1.9 *2*		+ *1*	29.6
PD1B		2.1 *2*	3.2 *3*	1.7 *2*		1.0 *1*	4.1 *4*		1.1 *1*				1.5 *2*	1.6 *2*			1.9 *2*		21.8
PD1C		2.1 *2*	3.5 *4*	1.8 *2*		1.1 *1*	4.3 *4*		1.1 *1*	0.5			1.8 *2*	2.8 *3*			2.0 *2*		7.8
PD1E		3.9 *4*	2.8 *3*	4.0 *6*	1.3 *1*		0.5	0.2	0.4	1.1 *1*	0.2	0.6	0.8 *1*	0.2		1.0 *1*		*1*	2.8

* Determined as *S*-carboxymethylcysteine.

† Peptide Sμ was determined to have a tryptophan residue by direct analysis for tryptophan in the methanesulphonic acid hydrolysate; peptide T15 was deduced to have a tryptophan residue because its strong A_{280} could not be due to any other residue; peptide PD1A probably has a tryptophan residue because its very strong A_{280} is unlikely to be due to its single tyrosine residue.

fractions containing both chains. The isolated chains, in volumes of about 100ml, were dialysed against each of several 10-litre volumes of 1% (v/v) acetic acid and stored frozen in that solution. The isolated light chain is very sparingly soluble in non-acidic conditions unless strong denaturants are added. The final yield of light chain was about 0.3mg/mg of immunoglobulin G subjected to sulphitolysis.

Summary of sequence analysis

The *N*-terminal sequence of the light chain was determined by Edman degradation (Smith, 1973*a*). The remainder of the sequence was established by Edman degradation and amino acid-composition analysis of peptides and peptide mixtures isolated from peptic, tryptic and CNBr digests of the light chain. Fig. 1 illustrates the fractionation of these three digests, Table 1 summarizes the amino acid compositions of peptic and tryptic peptides and peptide mixtures, and Fig. 2 outlines how the results of Edman degradations and amino acid-composition analyses were pieced together to deduce the amino acid sequence of the light chain. The methods used are described below.

CNBr digest

To 40mg of fully reduced and *S*-[^{14}C]carboxamidomethylated light chain (see under 'Complete reduction and alkylation of disulphide bonds') dissolved in 2ml of 70% (w/v) formic acid was added 140mg of CNBr; cleavage proceeded for 1h at room temperature, after which the digest was diluted with 20ml of water, freeze-dried and fractionated as shown in Fig. 1. Edman degradation of peptide mixture CNBr-I/II by the methods described below was reported previously (Smith, 1973*a*). Peptide CNBr-II was Edman-degraded after reaction with Braunitzer's reagent III, as described below.

Tryptic digest

A sample (19.4mg) of fully reduced and *S*-[^{14}C]-carboxymethylated light chain (see under 'Complete reduction and alkylation of disulphide bonds') was suspended in 1.9ml of 1% (w/v) NH_4HCO_3 and digested by adding 0.2mg of trypsin (EC 3.4.21.4; 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one-treated; Worthington Biochemical Corp., Freehold, NJ, U.S.A.) and incubating for 6h at 37°C. A copious gelatinous precipitate forms. Edman degradation of the precipitate showed it to be a mixture of long peptides with *N*-termini corresponding to the *N*-termini of almost all the variable-region tryptic peptides. In an attempt to circumvent this precipitation, I fully citraconylated (Atassi & Habeeb, 1972) another portion of *S*-[^{14}C]carboxymethylated light chain, digested it with trypsin as above, removed the citraconyl groups (Atassi & Habeeb, 1972), and redigested with trypsin as above. Unfortunately, although citraconylation did prevent precipitation, the gelatinous precipitate formed on removal of the citraconyl groups.

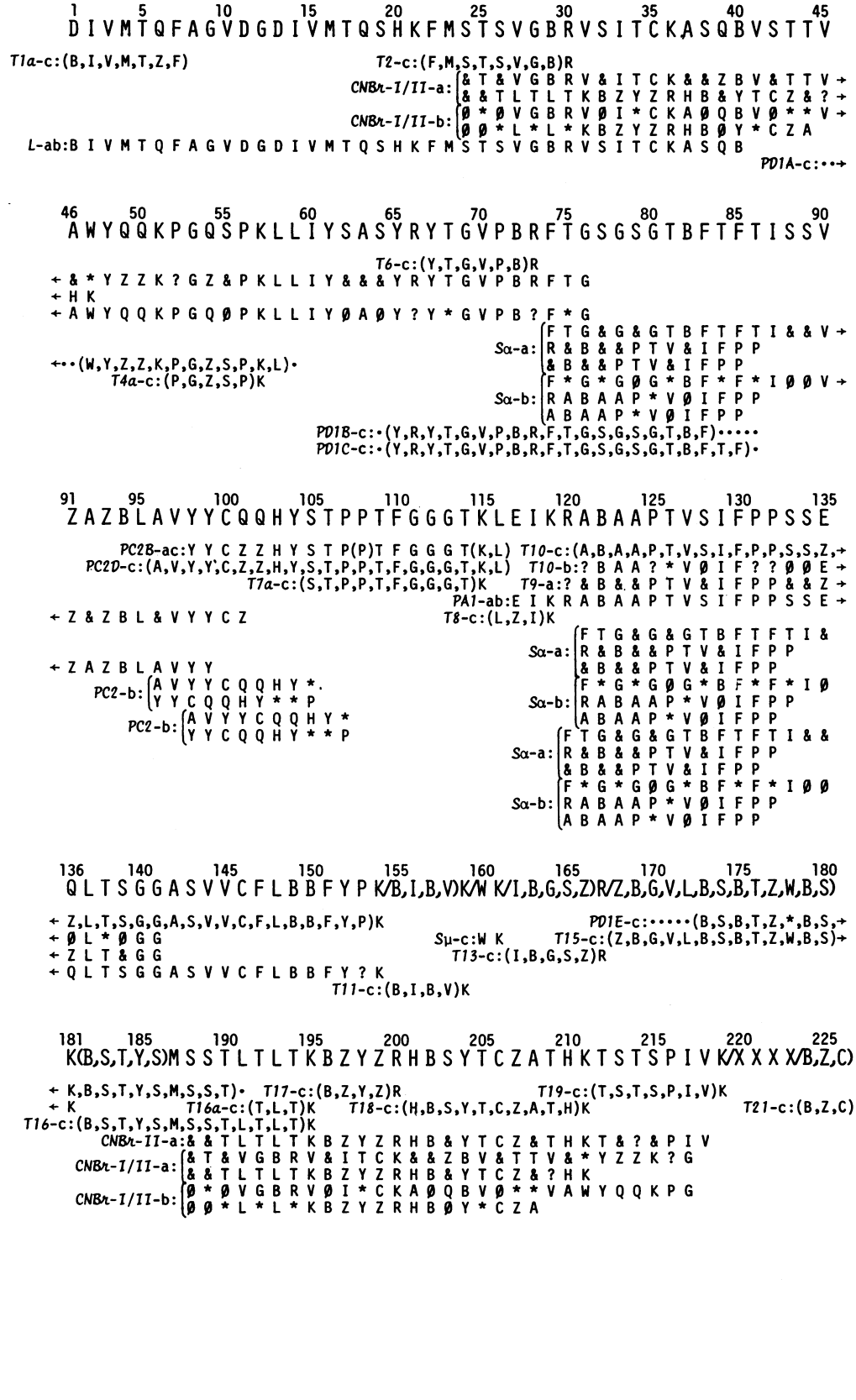
The tryptic digest was clarified by centrifugation and the supernatant was fractionated as shown in Fig. 1 (which gives the results for the portion that had undergone citraconylation), by a strategy based on the work of Svasti & Milstein (1972). The amino acid compositions of many of the peptides are shown in Table 1. Peptides T9 and T10, and peptide mixture Sα, all from a non-citraconylated preparation, were Edman-degraded after treatment with Braunitzer's reagent III (see under 'Edman degradations'); the results of these degradations are shown in Fig. 2.

Tryptic peptide T21 contains no lysine or arginine and is presumably the *C*-terminal sequence, as it is in other mouse κ-chains.

Three of the *S*-carboxymethylcysteine-containing tryptic peptides were isolated in a basic and an acidic form (T10 and T10′, T18 and T18′, and T21 and T21′), probably because of partial oxidation of the thioether group. Only one form each of the *S*-carboxymethylcysteine-containing peptides T3 (residues 31–36) and T9 were isolated.

The expected *S*-carboxymethylcysteine-containing tryptic peptide extending from residue 74 to residue 115 was never isolated in pure form, even though it was shown by Edman degradation to be present along with peptides T9 and T10 in at least one preparation of peptide mixture Sα. Possibly my failure to isolate it from other preparations of mixture Sα was due to nearly complete precipitation of the corresponding residues during those particular tryptic digestions.

The tryptic peptide Ser-Phe-Asn-Arg, corresponding to residues 220–223, was expected by homology to the constant-region sequences of other mouse κ-chains, but was never isolated. I expected this peptide to be present in peptide mixture Sγ and thus to have been recovered from the peptide 'map' (see Fig. 1). Possibly it reacted poorly with ninhydrin or ran off the end of the paper during the electrophoretic step of the peptide 'map'; alternatively, it might have been absent from peptide mixture Sγ because of adsorption on the polysaccharide matrix during Sephadex G-50 chromatography (see Fig. 1). Despite my failure to isolate this peptide, I think that the corresponding sequence is present in the myeloma MPC 11 κ-chain, since amino acid analysis of peptide mixture CBIII from the CNBr digest (see Fig. 1), which consists mostly of peptide CNBr-II, suggested the presence of at least some of these residues and was consistent with the presence of all of them (results not shown).


1 5 10 15 20 25 30 35 40 45
D I V M T Q F A G V D G D I V M T Q S H K F M S T S V G B R V S I T C K A S Q B V S T T V
T1a-c:(B,I,V,M,T,Z,F)
T2-c:(F,M,S,T,S,V,G,B)R
CNBr-I/II-a:
CNBr-I/II-b:
L-ab:B I V M T Q F A G V D G D I V M T Q S H K F M S T S V G B R V S I T C K A S Q B
PD1A-c:
46 50 55 60 65 70 75 80 85 90
A W Y Q Q K P G Q S P K L L I Y S A S Y R Y T G V P B R F T G S G S G T B F T F T I S S V
T6-c:(Y,T,G,V,P,B)R
T4a-c:(P,G,Z,S,P)K
Sα-a:
Sα-b:
PD1B-c:
PD1C-c:
91 95 100 105 110 115 120 125 130 135
Z A Z B L A V Y Y C Q Q H Y S T P P T F G G G T K L E I K R A B A A P T V S I F P P S S E
PC2B-ac:Y Y C Z Z H Y S T P(P)T F G G G T(K,L)
PC2D-c:(A,V,Y,Y,C,Z,Z,H,Y,S,T,P,P,T,F,G,G,G,T,K,L)
T7a-c:(S,T,P,P,T,F,G,G,G,T)K
T10-c:(A,B,A,A,P,T,V,S,I,F,P,P,S,S,Z,
T10-b:? B A A ? * V Ø I F ? ? Ø Ø E
T9-a:? & B & & P T V & I F P P & & Z
PA1-ab:E I K R A B A A P T V S I F P P S S E
T8-c:(L,Z,I)K
PC2-b:
PC2-b:
Sα-a:
Sα-b:
Sα-a:
Sα-b:
136 140 145 150 155 160 165 170 175 180
Q L T S G G A S V V C F L B B F Y P K/B,I,B,V)K/W K/I,B,G,S,Z)R/Z,B,G,V,L,B,S,B,T,Z,W,B,S)
PD1E-c:
Sμ-c:W K
T15-c:(Z,B,G,V,L,B,S,B,T,Z,W,B,S)
T13-c:(I,B,G,S,Z)R
T11-c:(B,I,B,V)K
181 185 190 195 200 205 210 215 220 225
K(B,S,T,Y,S)M S S T L T L T K B Z Y Z R H B S Y T C Z A T H K T S T S P I V K/X X X X/B,Z,C)
T17-c:(B,Z,Y,Z)R
T19-c:(T,S,T,S,P,I,V)K
T16a-c:(T,L,T)K
T18-c:(H,B,S,Y,T,C,Z,A,T,H)K
T21-c:(B,Z,C)
T16-c:(B,S,T,Y,S,M,S,S,T,L,T,L,T)K
CNBr-II-a:
CNBr-I/II-a:
CNBr-I/II-b:

Peptic digest

To 20ml of 1% (v/v) acetic acid containing 36mg of light chain, whose *C*-terminal cysteine residue was presumably sulphonated during chain separation (see above), but whose intrachain disulphide bonds were intact, was added 1.0ml of a 0.8mg/ml solution of pepsin (EC 3.4.23.1; Worthington) in water; the pH was 3.1. Digestion was allowed to proceed for 5.5h at room temperature. The pH was adjusted to 9 with NH_3 (the absence of any precipitation at this step indicating that virtually no light chain remained intact), and the digest was freeze-dried and fractionated as shown in Fig. 1. Each of the three major fractions isolated in the initial Sephadex G-100 step (see Fig. 1) was freed of non-volatile low-molecular-weight solutes by chromatography on Sephadex G-25 (fine grade) in 1% (v/v) acetic acid and then fully reduced and *S*-[^{14}C]carboxamidomethylated (see below) before further fractionation by chromatography on Sephadex G-25 (fine grade) as shown in Fig. 1.

Table 1 lists the amino acid compositions of several peptic peptides and peptide mixtures. Peptide PC2B and peptide mixtures PC2 and PA1 were Edman-degraded (PC2B and PC2 after first being treated with Braunitzer's reagent III), as described below; the results are shown in Fig. 2.

Complete reduction and alkylation of disulphide bonds

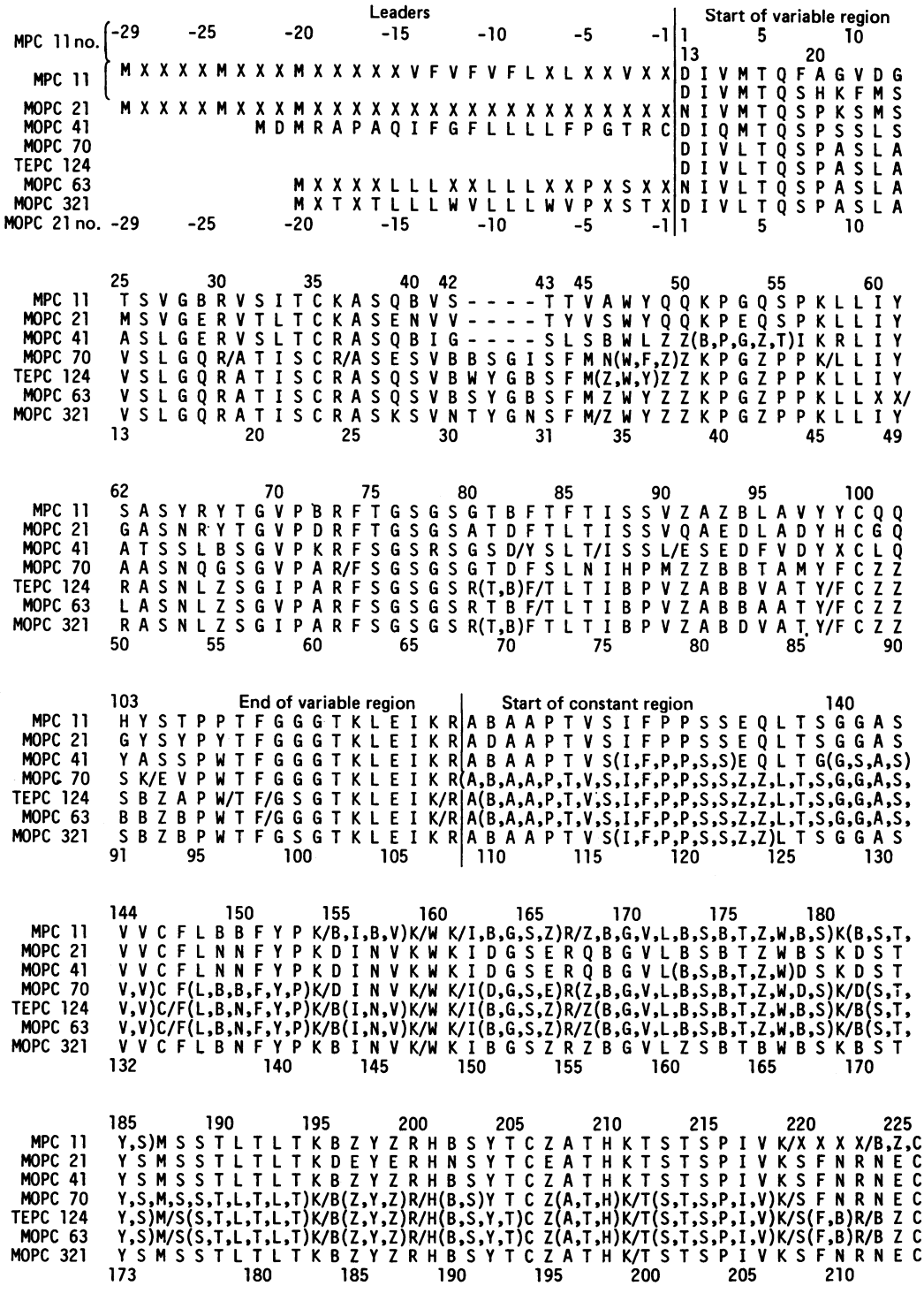
Protein or peptide was dissolved in 7M-guanidine hydrochloride/0.5M-Tris/0.01M-EDTA, pH adjusted to 8.3 with HCl. Dithiothreitol was added to give a concentration of 0.01M. After about 10min at room temperature, iodo[^{14}C]acetic acid or iodo[^{14}C]-acetamide (both from Amersham/Searle, Arlington Heights, IL, U.S.A.) was added to give a concentration of 0.022M and the sample incubated for 30min at room temperature. After alkylation with iodo[^{14}C]-acetic acid, the reduction and alkylation were repeated by adding dithiothreitol to 0.01M and iodoacetic acid to 0.1M to ensure complete reaction of thiol groups. Finally, 2-mercaptoethanol was added in about 5× molar excess of all the iodoacetic acid or iodoacetamide and the sample incubated for a further 30min at room temperature. The reduced and alkylated protein or peptide was freed of non-volatile solutes by dialysis against, or chromatography on Sephadex G-25 (fine grade) in, 1% (v/v) acetic acid, and then freeze-dried.

Ion-exchange chromatography: asymptotic-gradient elution

Each sample to be fractionated by ion-exchange chromatography was dissolved in or dialysed against the indicated initial buffer and applied to a column of the indicated Sephadex ion-exchange medium equilibrated with initial buffer. The inlet of the column was connected to the outlet of a closed mixing chamber containing initial buffer and a magnetic stirring bar; the inlet of this mixing chamber was connected initially to a reservoir of initial buffer. After the column had been washed with several column vol-

Fig. 2. *Summary of evidence for amino acid sequence of the myeloma MPC 11 full-length κ-chain*

The large upper-case letters give the deduced sequence, under which is given the evidence for that sequence. Peptides and peptide mixtures are identified in italics, followed by lower-case non-italicized letters designating the method of analysis: a and b, amino acid analysis of HI and NaOH hydrolysates respectively of amino acid thiazolinones from Edman degradations; c, amino acid-composition analysis (Table 1). Small upper-case non-italicized letters indicate sequence or composition of the indicated peptides or peptide mixtures. Results of Edman degradations of peptide mixtures are given as arrays as described in the text, one row of each array for each of the peptides presumed to be present in the mixture. Parentheses enclose and commas separate residues whose order was not determined. Solidus ('/') separates residues not shown to be connected to each other in the sequence; ')/', '/(', and ')/(' are shown simply as '/'. Dotted lines indicate possible extent of minor peptides in peptide mixtures. The single-letter abbreviations for the amino acids are as follows: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, unknown; Y, Tyr; Z, Glx. The following special symbols are used: &, Ser or Ala; *, no residue either identified or expected from deduced sequence; ?, residue expected from deduced sequence but not identified for technical reasons; Ø, no residue either identified or expected from deduced sequence, where the failure to identify it allows the distinction between Ser and Ala or between Gln and Glu (see the text). The following residues were identified by special considerations: residue 1 is shown as Asp rather than Asn solely because of the electrophoretic behaviour of the chain (Smith, 1973*a*); residue 42 was identified as Ser or Ala because the yield of Ala in the HI hydrolysate of step 19 of the Edman degradation of peptide mixture CNBr-I/II was much too high to be derived from residue 206 alone, and as Ser rather than Ala because of the absence of Ala from the corresponding NaOH hydrolysate; residue 83 was identified as Phe because the yield of Phe at step 10 of the Edman degradation of peptide mixture Sα was much too high for it to be derived solely from residue 130; residue 77 was identified as Ser or Ala because all other amino acids would have been positively identified by the results of Edman degradation of peptide mixture Sα, and as Ser rather than Ala from the absence of Ala from peptide mixtures PD1B and PD1C; residue 108 was not directly identified by Edman degradation of peptide PC2B because of technical difficulty but was deduced to be Pro because all other residues in peptides PC2B, PC2D and T7a had been placed in the sequence.


Leaders
Start of variable region
MPC 11 no.
MPC 11
MOPC 21
MOPC 41
MOPC 70
TEPC 124
MOPC 63
MOPC 321
End of variable region
Start of constant region

umes of initial buffer, asymptotic-gradient elution was started by connecting the inlet of the closed mixing chamber to a reservoir of the indicated final buffer.

Peptide 'map'

Tryptic peptide mixture Sγ, consisting of all the small tryptic peptides except those that were retarded by adsorption on the polysaccharide matrix during chromatography on Sephadex G-50 (fine grade) (see Fig. 1), was fractionated on Whatman 3MM paper as described by Bennett (1967). Descending chromatography for 17.5h in butan-1-ol/acetic acid/water/pyridine (15:3:12:10, by vol.) was followed by electrophoresis at right-angles at 2700V for 72min in water/acetic acid/pyridine (289:10:1, by vol.; pH3.6). Spots were detected by light spraying with ninhydrin (0.25mg/ml) in acetone, cut from the paper sheet, and washed with 7ml of water by suction filtration to elute the peptides.

Amino acid analyses

Peptides were dried into the bottoms of 10mm× 75mm test tubes and hydrolysed by the following modification (R. Niece, personal communication) of the method of Simpson *et al.* (1976). To each tube was added 20μl of 4M-methanesulphonic acid/12.5mM-3-(2-aminoethyl)indole. Many such tubes were placed inside a 16cm×25cm vacuum desiccator containing 18ml of 4M-H_2SO_4, and the desiccator was evacuated, sealed, clamped and heated at 120°C for 17–24h in an autoclave, as described by Smithies *et al.* (1971). After heating, the desiccator was cooled in a sink full of water, care being taken to loosen the clamp immediately lest the desiccator lid crack. The contents of each tube were partially neutralized with 0.18ml of 0.148M-trisodium citrate/2% (v/v) thiodiglycol. These neutralized solutions, or dilutions of them, were analysed with an automatic amino acid analyser.

Edman degradations

Edman degradations were performed on an automatic sequenator (Edman & Begg, 1967) made by Illinois Tool Works (Chicago, IL, U.S.A.), by the methods of Smithies *et al.* (1971). To counteract extraction of small peptides during the degradations, small peptides were treated with 2mg of 3-isothiocyanonaphthalene-1,5-disulphonic acid, disodium salt (Braunitzer's reagent III; Braunitzer *et al.*, 1971), in the reaction cup of the sequenator before degradation. Braunitzer's reagent III was loaded into the sequenator after the peptide had already been loaded and dried, rather than along with the peptide, since it forms an insoluble salt or aggregate with many peptides. Because of the low molar concentration of the reagent, the reaction was usually allowed to proceed for several hours before addition of phenyl isothiocyanate and continuation of the first step of the degradation.

The thiazolinone derivatives of the amino acids were identified by amino acid analysis of HI or NaOH hydrolysates as described by Smithies *et al.* (1971). The distinction between serine and alanine thiazolinones was made by comparing the yield of alanine in HI and NaOH hydrolysates: both give alanine on HI hydrolysis, but only alanine thiazolinone gives alanine on NaOH hydrolysis (Smithies *et al.*, 1971). The distinction between glutamine and glutamic acid thiazolinones was made in an analogous fashion: both give glutamic acid on HI hydrolysis, but only glutamic acid thiazolinone gives a high yield of glutamic acid on NaOH hydrolysis (Smithies *et al.*, 1971). Residues 11 and 13 were identified as aspartic acid rather than asparagine because there was no increase in the ammonia content of HI hydrolysates from those steps. *S*-[^{14}C]Carboxymethylcysteine or *S*-[^{14}C]carboxamidomethylcysteine residues were identified by radioactivity.

A somewhat unusual feature of this sequence determination was the use of Edman degradations of mixtures of peptides, so that several residues were liberated at each step. In Fig. 2 the results of these degradations are presented as arrays rather than as single sequences. Each array appears several times, aligned with the various sections of the sequence corresponding to the peptides present in the mixture. Separate arrays are given for the results of HI and

Fig. 3. *Comparison of mouse κ-chain sequences*

The one-letter abbreviations for amino acids and the meaning of punctuation marks are given in the legend to Fig. 2. Position numbers for the myeloma MPC 11 full-length κ-chain are shown at the top, and position numbers for the myeloma MOPC 21 κ-chain are shown at the bottom. The *N*-terminus of the myeloma MPC 11 full-length κ-chain is shown in two sections: residues 29 to 12 are on the top line and residues 13–24 are on the second line. The placement of the junction between the variable and constant regions is justified by Rose *et al.* (1977). Leader sequences are from the following sources: myeloma MPC 11 full-length κ, Rose *et al.* (1977); myeloma MOPC 21 κ, B. A. Kaplan, W. M. Kuehl & G. P. Smith (unpublished observations); κ-chains of myelomas MOPC 41, MOPC 63 and MOPC 321, Burstein & Schechter (1977). Sequences of residues 1 onwards of the κ-chains are from the following sources: myeloma MPC 11 full-length κ-chain, Smith (1973*a*) and present paper; myeloma MOPC 21 κ-chain, Svasti & Milstein (1972); κ-chains of myelomas MOPC 41 and MOPC 70, Gray *et al.* (1967); myeloma MOPC 321 κ-chain, McKean *et al.* (1973*a*); κ-chains of myelomas MOPC 63 and TEPC 124, McKean *et al.* (1973*b*).

NaOH hydrolysis. In most cases, all but one of the peptides present in the mixture were subsequently isolated and analysed separately; hence most residues of the remaining peptide could be unambiguously determined by difference. Residues that could not be unambiguously determined on this basis alone are discussed, along with other problematical residues, in the legend of Fig. 2. It should be noted too that the yields, rather than the mere presence, of the amino acids at each step provided a distinction between major and minor sequences even in the absence of other evidence about which of the residues liberated at one step is connected in the sequence to a given residue liberated at another step. In one of the peptide mixtures, PA1, a single major sequence so dominated over the minor sequences that it could be unambiguously determined on the basis of yields alone; it is that major sequence which is reported in Fig. 2.

Discussion

Structure of the myeloma MPC 11 full-length κ-chain

In Fig. 3 the sequence of the myeloma MPC 11 full-length κ-chain is compared with the sequences of other mouse myeloma κ-chains; the other myeloma MPC 11 κ-chain, which has no variable region and which is coded by a separate gene, is not included. Fig. 3 includes partial sequences (when known) of the *N*-terminal segments, called 'leaders' in this paper, that are present on the forms of these chains produced by translation of mRNA *in vitro* but absent from the forms of the chains found in or secreted by intact cells. The position numbers in the myeloma MOPC 21 κ-chain are given at the bottom, negative numbers being assigned to leader residues.

The secreted form of the myeloma MPC 11 full-length κ-chain has 12 extra residues (1–12) at its *N*-terminus. The first six residues (Asp-Ile-Val-Met-Thr-Gln; positions 1–6) of the extra 12-residue segment are shown immediately after the leader segment on the top line of Fig. 3, aligned with the obviously homologous positions 1–6 (Asn-Ile-Val-Met-Thr-Gln) of myeloma MOPC 21 κ-chain and with the homologous positions in other κ-chains, including residues 13–18 (Asp-Ile-Val-Met-Thr-Gln; second line of Fig. 3) of the myeloma MPC 11 full-length κ-chain itself. The second six residues of the extra 12-residue segment (Phe-Ala-Gly-Val-Asp-Gly; positions 7–12) could not be unambiguously aligned with any portion of mouse κ-chains. In Fig. 3 these six residues are shown at the end of the top line, aligned with positions 7–12 of myeloma MOPC 21 κ-chain. It should be noted that this alignment may not reflect the true evolutionary provenance of residues 7–12 of the myeloma MPC 11 full-length κ-chain, depending on how the gene for this chain arose (Smith, 1973*a*).

Other than the extra 12-residue *N*-terminal segment just discussed, no other abnormality was found in the entire myeloma MPC 11 full-length κ-chain, either in the variable region or in the constant region.

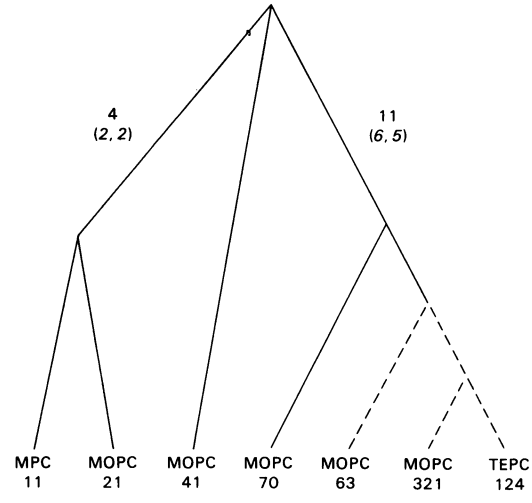


Fig. 4. *Evolutionary relationships among mouse κ-chain variable regions*

These were reconstructed by the method of minimum mutations as described in the text. The portion of the genealogy given in broken lines was assumed in advance, in accord with the results of McKean *et al.* (1973*b*). Positions 30–32 were ignored because of uncertainty in the alignment of the chains in this region. The non-italicized numbers assigned to connecting branches are mutational 'costs', which are explained in the text. The pairs of italicized numbers in parentheses are the corresponding mutational costs for residues 1–49 and residues 50–108 considered separately.

Variability of variable regions

Fig. 4 shows a genealogy relating the mouse κ-chain variable-region sequences shown in Fig. 3. This genealogy was reconstructed by the method of minimum mutations, in which all possible descents are examined in turn with the aim of finding the descent that requires the fewest mutations (Smith, 1973*b*, pp. 39–44). Five variable-region sequences were used: those of the κ-chains of mouse myelomas MPC 11 (positions 13–120), MOPC 21, MOPC 41, MOPC 70, and a common ancestral sequence for the κ-chains of mouse myelomas TEPC 124, MOPC 321 and MOPC 63. The ancestral sequence was reconstructed as described previously (Smith, 1973*b*, p. 40), assuming the descent reconstructed with the present methods by McKean *et al.* (1973*b*). The numbers assigned to the two connecting branches in Fig. 4, i.e. the two branches that do not terminate in one of the five sequences used in the reconstruction, are not the estimated number of mutations that occurred in

the corresponding ancestral sequences. Rather, each of these numbers, which I refer to as mutational 'costs', equals the extra mutations that would be required by the most parsimonious alternative genealogy in which the two groups into which the connecting branch divides the sequences are not preserved (Smith, 1973*b*, p. 43). For example, in the descent in Fig. 4 the left-hand connecting branch groups myeloma MOPC 21 κ-chain with myeloma MPC 11 κ-chain at one end of the branch, and groups all the other sequences together at the other end of the branch. This connecting branch is assigned a mutational cost of 4 because the most parsimonious alternative genealogy that does not group myeloma MOPC 21 κ-chain with myeloma MPC 11 κ-chain on the one hand and all the other sequences together on the other hand requires four more mutations than the overall best genealogy. These mutational costs are therefore a measure of the confidence with which the overall best genealogy can be said to be more like the true descent than the alternative genealogies. I do not know how to derive actual confidence limits from mutational costs, but in my opinion the mutational cost of 4 calculated for the branch separating myeloma MOPC 21 and myeloma MPC 11 κ-chain variable-region sequences from the other sequences is strong evidence that these two variable regions are, in fact, evolutionarily more closely related to each other than to the other sequences.

Descents were also reconstructed for the first and second halves of the variable regions considered separately; the numbers in parentheses in Fig. 4 are the resulting mutational costs for the two halves. The best descents for the two halves conformed to each other and to the best descent for the whole variable region. This conformity means that there is no evidence in these sequences for the switching of genealogical relationships in the middle of the variable region that might be expected from recombination among κ-chain variable-region genes (Smith, 1973*b*, pp. 81–84). A more critical residue-by-residue search by a previously described method (Smith, 1973*b*, pp. 81–84) also revealed no switching of genealogical relationships (results not shown). This supports the contention that, if somatic recombination among variable-region genes contributes to antibody diversity, it must be confined to recombinations between closely related variable-region genes, since recombinations between more distantly related genes would switch genealogical relationships and would thus be detectable (Smith, 1973*b*, pp. 81–84).

Constancy of constant regions

Extensive sequence data are available for the constant regions of seven different mouse κ-chains. As shown in Fig. 3, only two discrepancies among these sequences have been observed. In myeloma MOPC 41 κ-chain, the reported sequence at positions 127–132 (myeloma MOPC 21 κ-chain numbering system) is Gly(Gly,Ser,Ala,Ser)Val, whereas in the κ-chains of myelomas MOPC 21, MOPC 321 and MPC 11 the corresponding sequence is Ser-Gly-Gly-Ala-Ser-Val. The observed discrepancy is consistent with a switch in the myeloma MOPC 41 κ-chain of the serine residue at position 127 with one of the glycine residues at positions 128 and 129. The other discrepancy also involves an interchange of nearby residues. Residues 161–165 of myeloma MOPC 21 κ-chain are reported as Asx-Ser-Asx-Thr-Glx, whereas the corresponding residues of myeloma MOPC 321 κ-chain are reported as Glx-Ser-Asx-Thr-Asx. It seems likely that these interchanges of nearby residues are due to errors in the sequence determination rather than to real differences between the chains.

Part of this work was carried out in the laboratory of Dr. Oliver Smithies; I thank him for my training and for many helpful discussions. This work was supported by U.S. National Institutes of Health grants GM 20069 to Oliver Smithies and GM 22686 to myself, and by a Helen Hay Whitney postdoctoral fellowship to myself.

References

Atassi, M. Z. & Habeeb, A. F. S. A. (1972) *Methods Enzymol.* **25**, 546–553
Bennett, J. C. (1967) *Methods Enzymol.* **11**, 330–339
Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. & Petersen, U. (1971) *Hoppe-Seyler's Z. Physiol. Chem.* **352**, 1730–1732
Burstein, Y. & Schechter, I. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 716–720
Cole, R. D. (1967) *Methods Enzymol.* **11**, 206–208
Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–95
Gray, W. R., Dreyer, W. J. & Hood, L. (1967) *Science* **155**, 456–467
Kuehl, W. M. & Scharff, M. D. (1974) *J. Mol. Biol.* **89**, 409–421
Kuehl, W. M., Kaplan, B. A., Scharff, M. D., Nau, M., Honjo, T. & Leder, P. (1975) *Cell* **5**, 139–147
Laskov, R. & Scharff, M. D. (1970) *J. Exp. Med.* **131**, 515–541
McKean, D., Potter, M. & Hood, L. (1973*a*) *Biochemistry* **12**, 749–759
McKean, D., Potter, M. & Hood, L. (1973*b*) *Biochemistry* **12**, 760–771
Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. C. (1972) *Nature (London) New Biol.* **239**, 117–120
Rose, S. M., Kuehl, W. M. & Smith, G. P. (1977) *Cell* **12**, 453–462
Simpson, R. J., Neuberger, M. R. & Liu, T.-Y. (1976) *J. Biol. Chem.* **251**, 1936–1940
Smith, G. P. (1973*a*) *Science* **181**, 941–943
Smith, G. P. (1973*b*) *The Variation and Adaptive Expression of Antibodies*, Harvard University Press, Cambridge
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921
Svasti, J. & Milstein, C. (1972) *Biochem. J.* **128**, 427–444