Sequence Studies on the Heavy Chain of Rabbit Immunoglobulin A of different α -Locus Allotypes

By ALAN P. JOHNSTONE* and LAWRENCE E. MOLE†

Medical Research Council Immunochemistry Unit,
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX3 9U, U.K.

(Received 6 January 1977)

The amino acid sequence was determined of part of the variable region of heavy chain from rabbit immunoglobulin A of allotypes $\alpha 1$ and $\alpha 3$. Two corrections of the primary sequence of $\alpha 1$ γ -chains are reported; most of the structural correlates of the α -locus allotypes are confirmed. The amino acid sequence of the *N*-terminal 20 residues of α -negative molecules was also determined and found to be homologous to the human $V_H III$ subgroup. These molecules are present in a much higher proportion in the α -chain pool than in the γ -chain.

The Translocation Hypothesis (Dreyer & Bennett, 1965; Gally & Edelman, 1970) states that in the process of antibody synthesis any one of a number of immunoglobulin heavy-chain variable region genes is available for union with any of the class-specific constant-region genes. Whereas almost all structural data support this widely held theory, some data exist indicating a limitation of the permitted combinations. Thus an *N*-terminal heavy-chain sequence apparently confined to the IgM† class of human immunoglobulins has been reported (Bennett, 1968; Köhler, *et al.*, 1970). Also, Capra & Kehoe (1975*b*) showed an unequal distribution of the variable-region subgroups between the human heavy-chain classes. The aim of the present study was to investigate the Translocation Hypothesis by comparing the sequence of the heavy-chain variable region of rabbit IgA and IgG.

The structural correlates in the rabbit γ -chain variable region of the serologically defined α -locus allotypes, $\alpha 1$, $\alpha 2$ and $\alpha 3$ (Wilkinson, 1969; Mole *et al.*, 1971) have been challenged by Jaton *et al.* (1973), using data obtained from homogeneous antibodies. Because of the importance of these variable-region markers to theories concerning the

origin of antibody diversity, it was hoped that our present study would resolve these discrepancies.

The immunoglobulin pool of rabbit also contains molecules which lack the α -locus determinants (termed ' α -negative'). The *N*-terminus of these molecules is Glu-Glu-Gln (Prah *et al.*, 1973) and two isotopic variants of them (*x* and *y*) have been serologically defined (Kim & Dray, 1972); no additional information is available on these heavy chains.

The approach followed in the present study has been to determine the sequences of portions of the $\alpha 1$ and $\alpha 3$ α -chain variable region thought to be allotype-related (notably residues 1–20 and 80–85; Wilkinson, 1969; Mole *et al.*, 1971), and to compare these with those previously determined for γ -chain (Wilkinson, 1969; Mole *et al.*, 1971). The α -chain fragments used for these structural studies were $F(ab')_{2\alpha}$, Fd_α and the *N*-terminal CNBr 34-residue fragment (see Johnstone & Mole, 1977).

The present paper reports minor corrections of the $\alpha 1$ γ -chain sequence, but confirms the major α -locus allotypic correlates. The α -negative molecules were shown to be homologous to the human $V_H III$ subgroup, and were found in much higher proportions in the α -chain than in the γ -chain pool. The basic postulates of the Translocation Hypothesis are supported.

Experimental

Materials

Fragments $F(ab')_{2\alpha}$, Fd_α and the *N*-terminal 34-residue fragment of α -chain were prepared as described in the preceding paper (Johnstone & Mole, 1977). Peptides from the variable region of γ -chain were obtained by the methods of Wilkinson (1969) and Mole *et al.* (1971).

The 'V8-protease' enzyme (Houmard & Drapeau,

* Present address: Massachusetts General Hospital, Boston, MA 02114, U.S.A.

† To whom reprint requests should be addressed.

‡ Abbreviations: IgM, immunoglobulin M; IgA, immunoglobulin A; sIgA, secretory IgA; IgG, immunoglobulin G; $F(ab')_{2\alpha}$, peptic fragment of sIgA analogous to the $F(ab')_2$ fragment of IgG; Fd_α , *N*-terminal half of α -chain analogous to the *Fd* fragment of γ -chain; Glu , pyrrolid-2-one-5-carboxylic acid; CmCys, *S*-carboxymethylcysteine; V_{HI} , V_{HII} , V_{HIII} , heavy-chain variable-region subgroups I, II and III respectively; dansyl, 5-dimethylaminonaphthalene-1-sulphonyl; Dnp-, dinitrophenyl-.

1972) was a gift from Dr. G. R. Drapeau (Department of Microbiology, University of Montreal, Montreal, Que., Canada), who isolated it from *Staphylococcus aureus*, strain V8. Trypsin (Seravac Laboratories Ltd., Cape Town, S. Africa) was treated before use with L-1-chloro-4-phenyl-3-tosylaminobutan-2-one (TPCK; Kostka & Carpenter, 1964) to inhibit chymotryptic activity.

Thermolysin, α -chymotrypsin and carboxypeptidases A and B were purchased from Worthington Biochemical Corp. (Freehold, NJ, U.S.A.), pepsin was from Sigma Chemical Co. (St. Louis, MO, U.S.A.), and Pronase was from Kaken Chemical Co. (Tokyo, Japan).

Methods

Estimation of the N-terminal sequences of α - and γ -chains. The N-terminal 34-residue CNBr fragment of α - and γ -chains was dissolved in 0.5% (w/v) NH_4HCO_3 , pH 8.0 (250 nmol/ml), and digested with 'V8-protease' (100 μg of enzyme/ μmol of peptide) for 16 h at 37°C. Fragment F(ab')_{2a} and γ -chain were completely reduced and alkylated (O'Donnell *et al.*, 1970), excess reagents removed by dialysis against water and the proteins freeze-dried. The solid material was dissolved [F(ab')_{2a}] or suspended (γ -chain) in 0.5% (w/v) NH_4HCO_3 , pH 8.0 (10 mg/ml), and digested with 'V8-protease' at an enzyme/substrate ratio of 1:60 (w/w) for 16 h at 37°C.

In each case the digestion was terminated by freeze-drying and the whole digest subjected to one step of paper electrophoresis at pH 6.5 as described below. The N-terminal heavy-chain peptides, having no free amino group, were identified as such by their positive reaction with the hypochlorite/starch/iodide reagent of Pan & Dutcher (1956) and their lack of reaction with ninhydrin. These peptides were then eluted from the paper and their yields determined by amino acid analysis as described below.

Tryptic digestion of fragment Fd_a. Citraconic anhydride (BDH Chemicals, Poole, Dorset, U.K.) was used to block the ϵ -amino group of lysine residues (Dixon & Perham, 1968), as described by Mole *et al.* (1971). The [¹⁴C]carboxymethylated fragment Fd_a was dialysed into water, concentrated to 10 mg/ml and the pH adjusted to 8.0 with 1 M-NaOH. A 10–15-fold molar excess of citraconic anhydride over lysine was added in batches at room temperature (20°C), the pH being maintained at 7.5–8.5 by the addition of 1 M-NaOH. When the pH ceased to fall, the solution was dialysed against several changes of dilute NH_3 , pH 9.0, for 24 h.

The citraconylated fragment Fd_a was then digested with trypsin (1%, w/w) at pH 8.3 in the cell of a pH-stat at 37°C. After 2 h the digestion was stopped by freeze-drying, and the blocking group removed from lysine residues by suspending the protein in pyridine/

acetic acid/water (1:10:289, by vol.), pH 3.5, for 24 h at room temperature followed by freeze-drying. The digest was dissolved in formic acid (98%, v/v) and fractionated on a column (3.2 cm $\times$ 220 cm) of Sephadex G-50 (fine grade) in 50% (v/v) formic acid.

Enzymic digestion of peptides. The conditions of digestion were as follows: chymotryptic digestion of the N-terminal 34-residue fragment in 1% (w/v) NH_4HCO_3 , pH 8.2 (500 nmol of peptide/ml; 80 μg of enzyme/ μmol of peptide), at 37°C for 2 h (Aa3 peptide) or for 3.5 h (Aa1 peptide); tryptic digestion in 1% (w/v) NH_4HCO_3 , pH 8.2, at 37°C for Aa1 chymotryptic peptide C-1 (40 μg of enzyme/ μmol of peptide) for 4 h and for Aa1 peptide α 1-C-Ts4 (300 μg of enzyme/ μmol of peptide) for 2.5 h; peptic digestion of Aa3 N-terminal 34-residue fragment in 5% (v/v) formic acid, pH 2.0 (800 nmol of peptide/ml), at 37°C (80 μg of enzyme/ μmol of peptide) for 3 h; 'V8-protease' digestion in 0.5% (w/v) NH_4HCO_3 , pH 8.0, at 37°C for peptides 2CT-Ba and 2CT-Bb (35 μg of enzyme/ μmol of peptide) for 10 h and Aa1 peptide γ 1-Ts4 (90 μg of enzyme/ μmol of peptide) for 16 h; Pronase digestion of Aa1 tryptic peptide C-1-Ta in 1% (w/v) NH_4HCO_3 , pH 8.2 (500 nmol of peptide/ml), at 37°C (40 μg of enzyme/ μmol of peptide) for 3 h.

Analytical methods. Amino acid analysis, high-voltage paper electrophoresis and detection of peptides on paper by staining and radioautography were performed as detailed in Pratt & Mole (1975). Digestion with carboxypeptidases A and B was performed in 1% (w/v) NH_4HCO_3 , pH 8.2 (500 μg of enzyme/ μmol of peptide), at 37°C for 1–16 h; residues released were identified by amino acid analysis.

The dansyl-Edman procedure was used as described by Gray (1967) with modifications suggested by Hartley (1970). Dansyl-amino acids were resolved by t.l.c. on polyamide sheets (7.5 cm $\times$ 7.5 cm) (Woods & Wang, 1967), 0.3–1.0 nmol of the sample being applied to each side and a standard mixture to one side only. When the quantity of peptide allowed, the 2-anilino-5-thiazolinone derivative of the amino acid extracted in the butyl acetate wash was converted into the amino acid phenylthiohydantoin and identified by t.l.c. (Summers *et al.*, 1973). This allowed confirmation of the results obtained from the dansyl reaction, identification of some amino acids which the dansyl method could not detect (e.g. S-carboxymethyl-cysteine and tryptophan) and resolution of the acid and amide forms of aspartic acid and glutamic acid.

Automated sequence determination of peptides was performed on a Beckman 890C sequencer by using the programme of Brauer *et al.* (1975). The phenylthiohydantoins of amino acids were identified by g.l.c. (Hewlett-Packard 5830A), by t.l.c. (Summers *et al.*, 1973) and by back hydrolysis with HI as described by Smithies *et al.* (1971). In the t.l.c. system,

the threonine derivative was found close to the methionine derivative and not in the position described by Summers *et al.* (1973).

Characterization of radioactive *S*-carboxymethyl-cysteine-containing peptides in column eluates was carried out by tryptic-chymotryptic digestion followed by paper electrophoresis and radioautography as described by O'Donnell *et al.* (1970).

Notation for allotypes and peptides and the sequence numbering system throughout follows previous usage (Wilkinson, 1969; Mole *et al.*, 1971) as far as possible.

Results

Comparison of the *N*-terminal sequences of α- and γ-chains

Preliminary studies on the comparison of the α- and γ-chain *N*-termini have been reported by using Pronase digestion (Johnstone & Mole, 1975). By this method, the *N*-termini of α- and γ-chain of a given allotype appeared similar; however, the release by Pronase of several *N*-terminal peptides from one sequence (Wilkinson, 1969) complicated the purification of these peptides and rendered meaningful quantification difficult. To overcome this, the more specific enzyme 'V8-protease' (Houmard & Drapeau, 1972) was used and found to release one *N*-terminal peptide in good yield from each heavy-chain sequence. These peptides could be easily purified by one paper-electrophoresis step (Table 1).

The relative peptide yields obtained from the *N*-terminal 34-residue fragment agreed well with those from the whole γ-chain or fragment F(ab')_{2α} from which the CNBr fragment was derived (compare Tables 1*a* and 1*b*), showing that this fragment is representative of the whole heavy chain. The overall yields of *N*-terminal peptides from each chain were good (approx. 40% after paper electrophoresis), and this indicated a high degree of digestion and good recovery of the peptides from paper; the lower yield from whole γ-chain (approx. 20%) was attributed to insolubility of the substrate.

Only one peptide fraction with no free amino group (1-V8-1) was obtained from the Aa1 chain (Table 1). Its amino acid composition (Glu_{3.0}, Ser_{0.8}, Val_{0.7}, Leu_{0.4}) and charge of -3 (calculated from its mobility at pH6.5; Offord, 1966) indicated that this was a mixture of the peptides $\begin{smallmatrix} \text{Glu-Ser-Leu-Glu-Glu} \\ \text{Glu-Ser-Val-Glu-Glu} \end{smallmatrix}$ expected from the known γ-chain sequence (Wilkinson, 1969).

Three peptides with no free amino group were obtained from the Aa3 pool (Table 1). The composition of peptide 3-V8-2 (Glu_{3.0}, Ser_{0.8}, Leu_{1.0}) and its charge (-3) indicated that it was the *N*-terminal peptide $\begin{smallmatrix} \text{Glu-Ser-Leu-Glu-Glu} \end{smallmatrix}$ expected from the known γ-chain sequence. The composition of peptide 3-V8-1 was Leu_{1.0}, Glu_{5.0}, and its mobility indicated

Table 1. *N*-Terminal heavy-chain peptides obtained from a 'V8-protease' digest of α- and γ-chain *N*-terminal 34-residue fragment, F(ab')_{2α} and whole γ-chain. Mobilities are relative to pyrrolid-2-one-5-carboxylic acid. Yields of individual peptides are expressed as percentage of total *N*-terminal peptides recovered from each digest. Total yields of *N*-terminal peptides from each digest are expressed as mol of peptide/mol of polypeptide chain after one paper-electrophoresis step at pH6.5, no correction for losses being made.

Peptide fraction	Proposed structure*	Mobility		(a) Peptide yields from <i>N</i> -terminal CNBr fragments			(b) Peptide yields from F(ab') _{2α} and whole γ-chain				
		pH6.5	pH3.5	Aa1 α	Aa3 α	Aa3 γ	Aa1 α	Aa3 α	Aa1 γ	Aa3 γ	
1-V8-1	$\begin{smallmatrix} \text{Glu-Ser-Val-Glu-Glu} \\ \text{Glu-Ser-Leu-Glu-Glu} \end{smallmatrix}$	-0.90		65	—	—	60	—	70	—	
3-V8-2	$\begin{smallmatrix} \text{Glu-Ser-Leu-Glu-Glu} \\ \text{Glu-Ser-Leu-Glu-Glu} \end{smallmatrix}$	-0.90	-0.40	35	—	—	40	—	30	—	
3-V8-3	$\begin{smallmatrix} \text{Glu-Ser-Leu-Glu-Glu} \\ \text{Glu-Glu-Gln-Leu-Val-Glu} \end{smallmatrix}$	-0.90	-0.40	—	35.7	67.7	—	37.8	—	66.0	
3-V8-1	$\begin{smallmatrix} \text{Glu-Glu-Gln-Leu-Val-Glu} \\ \text{Glu-Glu-Gln-Leu-Glu-Glu} \end{smallmatrix}$	-0.84	-0.35	—	44.8 } 64.3	19.4 } 32.3	—	44.4 } 62.2	—	22.0 } 34.0	
		-0.97		—	19.5 }	12.9 }	—	17.8 }	—	12.0 }	
Total yields of <i>N</i> -terminal peptides recovered				0.42	0.35	0.39	0.43	0.35	0.18	0.24	

* The sequence of these peptides was incompletely determined (see the text for details).

a charge of -4 for a hexapeptide. After Pronase cleavage, the dansyl-Edman technique detected the sequence Leu-Glx-Glx; the third residue was identified as glutamate by the dansyl method without hydrolysis and was therefore C-terminal. Insufficient material remained for further characterization of the peptide, but the tentative sequence $\text{[Glu(Glu,Gln)Leu-Glu-Glu]}$ is proposed by homology with published γ -chain sequences (Wilkinson, 1969; Prahl *et al.*, 1973). The composition of peptide 3-V8-3 was $\text{Glu}_{4.0}, \text{Val}_{1.0}, \text{Leu}_{1.0}$ and its mobility indicated a charge of -3 for a hexapeptide. Thermolytic digestion produced quantitatively two products: the first (TL-1) had no free amino group, contained only glutamic acid on analysis and its mobility at both pH 3.5 and 6.5 was the same as that of [Glu-Glu-Gln] (Wilkinson, 1969); the second (TL-2) had a composition of $\text{Leu}_{0.9}, \text{Val}_{1.0}, \text{Glu}_{1.2}$, a charge of -1 and its partial sequence was determined by the dansyl-Edman method as Leu-Val-. Therefore the sequence of peptide 3-V8-3 is $\text{[Glu-(Glu,Gln)Leu-Val-Glu]}$. Peptides 3-V8-1 and 3-V8-3 are derived from α -negative molecules (Prahl *et al.*, 1973).

The yields of Aa1 peptides were the same from α - and γ -chains (Table 1), and agree well with the yield of Pronase peptides from γ -chain (Wilkinson, 1969). However, the *N*-terminal peptides derived from α -negative molecules (3-V8-1 and 3-V8-3) were re-

covered from Aa3 α -chain in a much higher relative yield than from Aa3 γ -chain (63% compared with 33%, Table 1). The yields of *N*-terminal peptides from Aa3 γ -chain agree closely with the Pronase-digestion data of Wilkinson (1969) in that the total amount of α -negative peptides recovered is approximately half of the amount of α -positive peptides; this rules out any possibility that the discrepancy between α - and γ -chain data is of a technical nature.

Sequence of Aa1 N-terminal 34-residue α -chain fragment

The products of the chymotryptic digest of this fragment were fractionated on a Sephadex G-50 column (1.80 cm $\times$ 220 cm) in 0.05 M-NH₃ and purified by paper electrophoresis (Table 2a). The *N*-terminal peptide obtained (C-1) was further digested with trypsin and the products were isolated by paper electrophoresis (Table 2b). The *N*-terminal peptide obtained from this digest (C-1-Ta) was further digested with Pronase and two products were isolated (Table 2c). The sequence obtained by characterization of these peptides (Fig. 1) agreed with that of the Aa1 γ -chain, except for residues 15–17, which are rearranged (Wilkinson, 1969). To check this discrepancy, the tryptic-chymotryptic peptide C-1-Tb was isolated from γ -chain and shown, by the

Table 2. Isolation and amino acid compositions of the major chymotryptic, tryptic and Pronase peptides of Aa1 *N*-terminal 34-residue fragment

Compositions are given as mol of residues/mol of peptide. Yields are expressed as mol of peptide recovered after one paper-electrophoresis step per mol of original peptide digested, no correction for losses being made. Mobilities at pH 6.5 are relative to aspartic acid (-1.0) and ϵ -Dnp-lysine (0), at pH 3.5 are relative to lysine ($+1.0$), pyrrolid-2-one-5-carboxylic acid (-1.0) and ϵ -Dnp-lysine (0), and at pH 1.9 are relative to lysine ($+1.0$) and ϵ -Dnp-lysine (0)

	(a) Chymotryptic peptides of <i>N</i> -terminal CNBr fragment			(b) Tryptic peptides of peptide C-1		(c) Pronase peptides of tryptic peptide C-1-Ta	
	C-1	D-1	E-1	C-1-Ta	C-1-Tb	C-1-Ta-Pr1	C-1-Ta-Pr5
Arg	1.0	—	—	1.0	—	—	0.8
CmCys	—	0.7	—	—	—	—	—
Asp	0.2	—	0.4	—	—	—	—
Thr	1.9	2.5	0.6	—	1.8	—	—
Ser	2.1	1.3	2.1	2.0	—	0.9	1.0
Glu	2.9	—	—	3.5	—	2.3	1.1
Pro	1.8	—	—	—	2.0	—	—
Gly	3.2	1.2	0.6	2.1	1.2	—	2.0
Ala	0.2	0.5	—	—	—	—	—
Val	1.5	0.5	—	0.6	1.1	0.65	—
Leu	2.3	1.0	1.0	0.4	2.0	0.35	—
Tyr	—	—	0.8	—	—	—	—
Phe	—	1.0	—	—	—	—	—
Mobility pH 6.5	-0.37	-0.30	—	—	—	—	—
pH 3.5	—	—	—	0.0	$+0.20$	-0.37	$+0.48$
pH 1.9	—	—	$+0.06$	—	—	—	—
Ninhydrin colour	None	Blue	Yellow	None	Blue	None	Blue
Yield	0.50	0.29	0.20	0.60	0.41	0.55	0.40

Fig. 1. Amino acid sequence of residues 2-32 of Aa1 heavy-chain N-terminal fragment
Sequence determination from the N-terminus was by the dansyl-Edman technique (see under 'Analytical methods'), indicated by →.

Table 3. Isolation and amino acid compositions of the major chymotryptic peptides of Aa3 α- and γ-chain N-terminal 34-residue fragment

Compositions are given as mol of residues/mol of peptide. Mobilities and yields are expressed as in Table 2.					
	(a) α -Chain			(b) γ -Chain (N-terminal peptides only)	
	2CT-Ba	2CT-Bb	2CT-Dc	γ CT-A-2	γ CT-A-1
Lys	0.7	—	—	0.7	0.3
CmCys	—	—	0.8	—	—
Asp	1.4	0.8	0.5	1.0	0.4
Thr	—	1.1	1.8	—	0.7
Ser	2.5	2.0	1.3	2.2	2.1
Glu	3.2	5.2	—	3.2	5.9
Pro	1.4	1.2	—	0.9	1.0
Gly	3.0	3.5	1.3	3.2	4.2
Ala	1.0	—	1.0	1.0	—
Val	1.3	1.5	—	1.0	1.4
Ile	0.6	—	—	—	—
Leu	2.6	3.4	—	2.9	3.8
Phe	—	—	1.1	—	—
Mobility at pH6.5	-0.58	-0.70	-0.36	-0.59	-0.67
Ninhydrin colour	Blue	None	Blue	Blue	None
Yields	0.11	0.17	0.23	0.25	0.10
	0.28			0.35	

dansyl-Edman technique, to have the same amino acid sequence as found in α -chain (Fig. 1).

Sequence of Aa3 N-terminal 34-residue α -chain fragment

Two *N*-terminal peptides were obtained from the chymotryptic digest of the Aa3 fragment by gel filtration and paper electrophoresis (2CT-Ba and 2CT-Bb, Table 3a). These were characterized by digestion with 'V8-protease' (Tables 4a and 4b). The structure of peptide 2CT-Ba (Fig. 2) agreed with the sequence of Aa3 γ -chain; peptic digestion of the Aa3 *N*-terminal 34-residue α -chain fragment confirmed this sequence (Fig. 2). A minor variant of the cysteine-22-containing chymotryptic peptide (2CT-Da-II; isolated in 20% yield relative to peptide 2CT-Dc) was found in α - and γ -chain (Fig. 2).

The second *N*-terminal chymotryptic peptide obtained (2CT-Bb) was also observed by Wilkinson (1969) in γ -chain (peptide 'C1a'); however, in the earlier study, it was not isolated in sufficient amounts for complete characterization, and was deduced to be the *N*-terminal 11-residue peptide of Aa3 chains produced by a partial chymotryptic cleavage. Our study (Table 4b and Fig. 3) showed that it is in fact the *N*-terminal 20-residue fragment of *a*-negative molecules and has the two *N*-terminal sequences previously detected in the whole Aa3 α - and γ -chain

pool (see above). To confirm that this peptide was not confined to α -chain, the structure of the γ -chain *N*-terminal 34-residue fragment was re-investigated (Tables 3b and 4c) and found to be in complete agreement with the data from α -chain. This 20-residue sequence of *a*-negative molecules has a remarkably high degree of homology (80%) with that of the human V_H III subgroup (Fig. 3).

The yields of the *N*-terminal chymotryptic peptides (Tables 3a and 3b) show that the *a*-negative chains comprise 61% of the Aa3 α -chain pool but only 29% of the Aa3 γ -chain pool; these values are in close agreement with those obtained from studies on whole F(ab')₂ α and γ -chain (62% and 34% respectively; Table 1b).

Sequence studies on Aa1 fragment Fd_α

The tryptic digest of citraconylated Aa1 fragment Fd_α gave the elution profile on Sephadex G-50 shown in Fig. 4. The first two peaks (TC_A and TC_B) contained [¹⁴C]carboxymethylcysteine-containing peptides not found in γ -chain (O'Donnell *et al.*, 1970) and these were assumed to derive from the α -chain constant region. The peptides characteristic of cysteine-22 and cysteine-92 were found in the third peak (pool A) and its trailing edge (pool B).

Pool A was further fractionated on a Sephadex G-50 column (3.2 cm × 220 cm) in 0.05 M-NH₃ and

Table 4. Isolation and amino acid composition of the major 'V8-protease' peptides of the *N*-terminal α - and γ -chain chymotryptic peptides

Compositions are given as mol of residues/mol of peptide. Mobilities and yields are expressed as in Table 2. The α -chain equivalent of the minor variant *N*-terminal γ -chain peptide (γ CT-A-1-e) was not isolated, because its existence was not realized at the time of this α -chain study; however, this peptide was isolated from fragment F(ab')₂ α in a separate experiment (see Table 1; peptide 3-V8-1).

	(a) α -Chain peptide 2CT-Ba		(b) α -Chain peptide 2CT-Bb			(c) γ -Chain peptide γ CT-A-1			
	2CT-Ba-I	2CT-Ba-III	2CT-Bb-II	2CT-Bb-III	2CT-Bb-IV	γ CT-A-1-e	γ CT-A-1-c	γ CT-A-1-b	γ CT-A-1-a-I
Lys	—	0.8	—	—	—	—	—	—	—
Asp	—	1.0	—	—	—	—	—	0.3	—
Thr	—	—	—	—	0.9	—	—	—	0.8
Ser	0.7	1.4	—	0.9	1.0	—	—	0.8	0.9
Glu	2.9	—	3.3	1.9	—	5.0	4.1	1.9	—
Pro	—	1.0	—	1.0	—	—	—	0.9	—
Gly	—	3.1	—	3.0	1.4	—	—	3.0	1.1
Ala	—	1.0	—	—	—	—	—	—	—
Val	—	1.0	1.0	1.0	—	—	1.0	1.0	—
Leu	1.0	2.0	1.0	1.0	2.2	1.0	1.0	1.1	2.0
Mobility pH3.5	−0.42	+0.32	−0.34	+0.13	+0.26	—	—	—	—
pH6.5	−0.92	—	−0.91	−0.31	—	−1.07	−0.92	−0.29	—
pH1.9	—	—	—	—	—	—	—	—	+0.15
Ninhydrin colour	None	Blue	None	Blue	Yellow	None	None	Blue	Yellow
Yield	0.36	0.36	0.46	0.42	0.37	0.125	0.34	0.50	0.50
						0.465

Fig. 2. Amino acid sequence of residues 2-27 of Aa3 α -chain N-terminal fragment

Sequence determination from the N-terminus was by the dansyl-Edman technique, indicated by $\rightarrow$, and from the C-terminus by hydrolysis with carboxypeptidase A, indicated by $\leftarrow$.

Fig. 3. Amino acid sequence of α -negative N-terminal chymotryptic peptide 2CT-Bb. Sequence determination from the N-terminus was by the dansyl-Edman technique, indicated by $\rightarrow$, and from the C-terminus by hydrolysis with carboxypeptidase A, indicated by $\leftarrow$. The human sequence is taken from Capra & Kehoe (1975a).

Fig. 4. Fractionation of a tryptic digest of citraconylated Aa1 Fd_α fragment
Gel filtration was conducted on a column (3.2cm × 220cm) of Sephadex G-50 in 50% (v/v) formic acid. Fraction size was 15ml. —, A₂₈₀; ---, radioactivity (c.p.m./ml of eluate).

Table 5. Isolation and amino acid compositions of tryptic peptides of Aa1 peptide C-Ts4
Compositions are given as mol of residues/mol of peptide. Mobilities and yields are expressed as in Table 2.

	(a) α-Chain			(b) γ-Chain γ1-Ts4
	α1-C-Ts4-Da	α1-C-Ts4-Ca	α1-C-Ts4-Ba	
Lys	0.9	1.0	—	—
Arg	—	—	0.9	0.6
CmCys	—	—	0.9	0.8
Asp	—	1.1	1.0	1.0
Thr	1.0	2.9	4.2	4.2
Ser	1.1	1.3	1.0	1.2
Glu	—	—	1.1	1.1
Pro	—	—	1.1	1.1
Ala	—	—	1.8	2.0
Val	—	1.0	—	—
Ile	1.0	—	0.6	0.8
Leu	—	1.1	0.3	0.2
Tyr	—	—	0.9	0.9
Phe	0.9	—	0.8	0.9
Mobility	—	—	—	—
pH6.5	—	—	-0.32	-0.35
pH3.5	+0.58	+0.40	—	—
Yield	0.32	0.41	0.55	—

Fig. 5. Amino acid sequence of Aa1 α-chain peptide α1-C-Ts4
Sequence determination from the N-terminus was by automated sequence analysis, indicated by →, or by the dansyl-Edman technique, indicated by ←, and from the C-terminus by hydrolysis with carboxypeptidase B, indicated by →, or with carboxypeptidases A and B, indicated by ←. (X) indicates unidentified residue on automated sequence analysis.

the peptide spanning residues 67–94 (α 1-C-Ts4) was resolved from the two other variable-region peptides in pool A (residues 11–38 and 39–66) by this means, as described for γ -chain peptides (Mole *et al.*, 1971). The sequence of peptide α 1-C-Ts4 was determined by automated sequence analysis and by characterization of its tryptic peptides (Fig. 5 and Table 5a). The sequence of this peptide contained an extra threonine residue at position 84 compared with that reported for γ -chain (O'Donnell *et al.*, 1970; Mole *et al.*, 1971). To investigate this discrepancy, the corresponding γ -chain peptide was isolated (Table 5b) and its sequence, determined by the dansyl-Edman method, shown to be identical with that found for α -chain. To confirm this corrected sequence the γ -chain peptide (γ 1-Ts4, Table 5b) was digested with 'V8-protease' and two products were isolated in good yield by paper electrophoresis at pH 6.5 (V8-1 and V8-2; Fig. 5). The amino acid composition (Thr_{2.7}, Ser_{1.1}, Glu_{1.2}, Pro_{0.9}, Ile_{0.9}, Leu_{0.2}) and sequence (Fig. 5) of peptide V8-1 supported this correction.

Sequence studies on Aa3 fragment Fd_a

The tryptic digest of citraconylated Aa3 fragment Fd_a gave a similar profile to that obtained from the Aa1 fragment (Fig. 4). The distribution of constant- and variable-region [¹⁴C]carboxymethylcysteine residues across this profile was also similar to that in the Aa1 digest.

The tryptic peptide spanning residues 67–94, present in pool A, was not purified sufficiently for sequence analysis. The peptide spanning residues 81–94 of the Aa3 α -chain, released by CNBr cleavage of the partial methionine residue (i.e. methionine only present in some molecules of the heterogeneous pool) at position 80, was isolated from pool B (Fig. 4) by gel filtration on a Sephadex G-50 column (1.8 cm $\times$ 220 cm) in 0.05 M-NH₃, followed by paper electrophoresis at pH 6.5 and 8.9 (peptide α 3-M-Ts4-A-b; Table 6). A second form of this peptide was obtained in lower yield (20% relative to the major yield peptide) on electrophoresis at pH 6.5 (mobility –0.39, relative to aspartic acid –1 and ϵ -Dnp-lysine 0); this corresponds to the *N*-formylated peptide observed in γ -chain (Mole *et al.*, 1971) and later characterized by Mole (1975). The sequence of the unmodified peptide (α 3-M-Ts4-A-b; Table 6) was completely determined, by the dansyl-Edman method and digestion with carboxypeptidases A and B, and found to be identical with that of γ -chain (Mole *et al.*, 1971): Thr-Ser-Leu-Thr-Ala-Ala-Asp-Thr-Ala-Thr-Tyr-Phe-CmCys-Ala-Arg.

A second radioactive peptide (α 3-M-Ts4-A-a) was isolated together with peptide α 3-M-Ts4-A-b and was only resolved from it by electrophoresis at pH 8.9 (Table 6). Its sequence was completely determined, by the dansyl-Edman method and digestion with carboxypeptidases A and B, as: Pro-Ala-Pro-Glu-Gln-

Table 6. Isolation and amino acid compositions of peptides obtained from pool B (see Fig. 4) of the tryptic digest of Aa3 fragment Fd_a

Compositions are given as mol of residues/mol of peptide. Mobilities are relative to aspartic acid (–1.0) and ϵ -Dnp-lysine (0). Data for the γ -chain peptide are taken from Mole *et al.* (1971).

	α 3-M-Ts4-A-a		α 3-M-Ts4-A-b	Aa3 γ Ts4
	22 h hydrolysis	72 h hydrolysis		
Arg	0.6	0.9	0.7	1.0
CmCys	0.7	1.0	0.9	+ve
Asp	0.3	—	1.2	1.1
Thr	1.2	1.2	3.5	3.9
Ser	0.3	—	1.3	1.1
Glu	1.6	2.0	0.2	0.3
Pro	1.8	1.6	—	—
Gly	1.1	1.5	0.3	0.3
Ala	1.3	1.5	3.8	3.9
Val	2.1	2.5	0.3	0.2
Ile	0.8	0.9	—	—
Leu	1.0	1.1	1.2	1.0
Tyr	—	—	1.0	1.0
Phe	—	—	0.9	1.0
Mobility				
pH 6.5	–0.18		–0.18	–0.19
pH 8.9	–0.45		–0.38	—
Sp. radio-activity (c.p.m./nmol)	2.2 $\times$ 10 ³		2.4 $\times$ 10 ³	

Thr-Val-Val-Val-Gly-CmCys-Leu-Ile-Arg. This peptide was also detected, but not purified, in the digest of Aa1 Fd_a, and it presumably derives from the α -chain constant region. The similar values obtained for the specific radioactivity (c.p.m./nmol) of this peptide and α 3-M-Ts4-A-b (Table 6) indicate that the former is involved in an intrachain disulphide bridge.

Sequence of Aa1 γ -chains with methionine at position 80

The structure of the peptide spanning residues 81–94 of Aa1 γ -chain, which is released by CNBr cleavage of the partial methionine residue at position 80, was investigated in attempts to resolve the discrepancy between the sequence data obtained from the Aa1 pool (only 20% of which has the methionine variant) and from homogeneous antibodies with methionine at position 80 (Mole *et al.*, 1971; Jaton *et al.*, 1973). Two alternative structures for this region of Aa1 γ -chain, linked to the presence or absence of methionine at position 80, have been detected by peptide 'mapping' (Mole, 1975).

The peptide from molecules with the methionine variant was isolated from a tryptic digest of citraconylated γ -chain, by gel filtration (Mole *et al.*, 1971),

Table 7. Isolation and amino acid compositions of cysteine-92-containing peptides from Aa1 γ -chains with methionine at position 80

Compositions are given as mol of residues/mol of peptide. Mobilities at pH 6.5 are expressed as in Table 2; mobilities on chromatography (see the text) are expressed relative to ϵ -Dnp-lysine (1.0) and the origin (0)

	1A1-A	1A1-B
Arg	1.0	1.0
CmCys	+ve	+ve
Asp	1.1	1.2
Thr	4.6	4.5
Ser	1.2	1.2
Glu	1.2	1.3
Pro	1.2	—
Gly	0.4	0.7
Ala	2.1	2.2
Leu	0.5	0.8
Tyr	0.9	0.9
Phe	0.9	1.0
Mobilities pH6.5	-0.38	-0.38
chromatography	0.12	0.20

paper electrophoresis at pH 6.5 and paper chromatography in the solvent pyridine/pentan-1-ol/water (9:9:7, by vol.) adjusted to pH 7.0 with acetic acid. Two short (15 residues long) cysteine-92-containing peptides were resolved by this method (Table 7).

The dansyl-Edman technique gave the *N*-terminal sequence of peptide 1A1-A as Thr-Ser-Pro-Thr-Thr and that of peptide 1A1-B as: Thr-Ser-Leu-Thr-Thr-Glx-Asx-. The partial sequence of the methionine variant molecules is therefore Thr-Ser-^{Leu}_{Pro}-Thr-Thr- (Glu, Asp, Thr, Ala, Thr, Tyr, Phe, Cys, Ala, Arg).

There is therefore a partial linkage between methionine at position 80 and leucine at position 83; otherwise, the two sequences are identical (Fig. 6). The sequence correction in this region (see above) places the methionine residue at position 80 and not as a variant of lysine at position 79. Consequently, CNBr cleavage and tryptic digestion release two peptides containing cysteine-92 which differ in length by one amino acid. This is the basis of the two spots observed by Mole (1975) in peptide 'maps' of Aa1 γ -chain pool.

Discussion

'V8-protease' (Houmard & Drapeau, 1972), under the conditions used, cleaved only on the carboxyl side of glutamic acid residues. Furthermore the peptide bond on the *N*-terminal side of glutamine or glutamic acid was resistant to cleavage (peptides 1-V8-1, 3-V8-1, 3-V8-2 and 3-V8-3; Table 1), whereas

Aa1 peptide γ 1-Ts4 was quantitatively cleaved between glutamic and aspartic residues (Fig. 5).

The two mistakes in the earlier pooled data for Aa1 γ -chain (Wilkinson, 1969; Mole *et al.*, 1971), corrected in the present study, can be attributed to the indirect methods of sequence determination used in the regions in question. The incorrect sequence for heavy-chain residues 15-17 has also been reported for a homogeneous antibody (Jaton, 1974); however, the sequence of several other antibodies supports the corrected sequence (van Hoergarden & Strosberg, 1976; L. E. Cannon & M. N. Margolies, personal communication).

The isolation of a peptide bearing a3 allotypic determinants has been reported (Ansari *et al.*, 1976); however, its amino acid sequence, in the region of cysteine-22, does not agree with that of a3 pooled heavy chain (Wilkinson, 1969) confirmed in our present study. Most of the structural data presented by Ansari *et al.* (1976) on their 'cysteine-22' peptide are consistent with the sequence around cysteine-425 of the γ -chain (Hill *et al.*, 1967).

Most of the *a*-locus allotypic structural correlates are confirmed (Fig. 6); however, the correction at position 84 increases the homology between the three allotypes in this region, and only residues 85 and 86 now appear to be allotype-related. The variations in these two residues are the basis of the chemically defined variable-region marker described by Mole (1975). A partial linkage of leucine at position 83 with methionine at position 80 was found in Aa1 molecules, but otherwise the sequence in this region is homogeneous. Variations in the region 28-33 (Fig. 6) are probably attributable to heterogeneity in the first hypervariable region; hence, these residues are not classed as allotype-related.

The Aa1, Aa2 and Aa3 sequences are each more homologous to the human V_HII subgroup than to the V_HI or V_HIII subgroups (Mole, 1975; Capra & Kehoe, 1975a; Jaton, 1974, 1975). However, the sequence of the *a*-negative molecules obtained in the present study correlates well with the prototype human V_HIII subgroup in that 16 out of 20 residues are identical (Fig. 3). This indicates that the *a*-negative and *a*-positive molecules are encoded at different loci, in agreement with serological studies (Kim & Dray, 1972). The relevance of the heterogeneity at position 5 in the *a*-negative molecules to the serologically defined variants (*x* and *y*) is unknown. The remarkably high degree of homology between the rabbit *a*-negative chains and the human V_HIII subgroup agrees with studies of Capra *et al.* (1973), who found unblocked V_HIII molecules, to a varying extent, in the γ -chain pool of many mammals. These workers studied only heavy chains with a free *N*-terminus and so did not observe the rabbit V_HIII sequence.

The *a*-negative molecules (V_HIII subgroup) are

	1	10	20
Aa1	<u>Glu-Ser-Val</u> Leu	<u>Glu-Glu-Ser-Gly</u> Arg	<u>Glu-Val-Thr</u> Pro-Gly-Thr-Pro-Leu-Thr-Leu-
Aa2	<u>Glu-Ser-Val</u> Lys	<u>Glu-Ser-Glu-Gly</u> Gly	<u>Glu-Phe-Lys</u> Pro-Thr-Asp-Thr-Leu-Thr-Leu-
Aa3	<u>Glu-Ser-Leu</u> Glu	<u>Glu-Ser-Gly</u> Gly	<u>Glu-Val-Lys</u> Pro-Gly-Ala-Ser-Leu-Thr-Leu-
α -negative	<u>Glu-Glu-Gln-Leu</u> Val-Glu	<u>Glu-Ser-Gly-Gly-Leu-Val</u> Gln-Pro-Glu-Gly-Ser-Leu-Thr-Leu	
Aa1	<u>Thr-Cys-Thr</u> Val-Ala	<u>Val-Ser-Gly-Phe-Ser-Leu-Ser-Ser-Tyr</u> Ala-Met-Gly-Trp-Val-Arg	38
Aa2	<u>Thr-Cys-Thr</u> Val-Ala	<u>Val-Ser-Gly-Phe-Val-Ser-Ser-Tyr</u> Met-Gly-Asp-Ser-Trp-Val-Arg	
Aa3	<u>Thr-Cys-Thr</u> Lys	<u>Thr-Als-Ser-Gly-Phe-Asn</u> Gly-Ser-Ser-Phe-Tyr-Met-Tyr-	
Aa1	<u>Gly-Arg-Phe</u> Thr-Ile	<u>Ser-Lys-Thr-Ser-Thr-Thr-Val-Asp-Leu-Lys-Leu-Thr-Ser-Pro</u> Ile-Met-Leu	80
Aa2	<u>Ser-Thr-Ile</u> Thr-Arg	<u>Ser-Thr-Ile</u> Thr-Arg	(-)-Thr-Ser-Leu-Met
Aa3	<u>Gly-Arg-Phe</u> Thr-Ile	<u>Ser-Lys</u> Ser-Lys	(-)-Thr-Ser-Leu-Met
Aa1	<u>Thr-Thr-Glu</u> Asp-Thr-Ala-Thr-Tyr-Phe-Cys-Ala-Arg	90	94
Aa2	<u>Thr-Ala-Ala</u> Asp-Thr-Ala-Thr-Tyr-Phe-Cys-Ala-Arg		
Aa3	<u>Thr-Ala-Ala</u> Asp-Thr-Ala-Thr-Tyr-Phe-Cys-Ala-Arg		

Fig. 6. Comparison of the Aa1, Aa2, Aa3 and α -negative variable-region amino acid sequences from pooled heavy chain. Residues in boxes are those thought to be related to the α -locus allotype. V indicates a large amount of heterogeneity at this position, and (-) an unidentified variant. Sequences underlined were found in α - as well as γ -chain. Aa2 data are from Mole (1975, 1976) and A. N. Barclay & L. E. Mole (unpublished work); Aa1, Aa3 and α -negative data are from Wilkinson (1969), Mole *et al.* (1971) and this study.

present in the α -chain pool at approximately twice the amount in γ -chain (in Aa3 rabbits). The α -negative molecules in Aa1 γ -chain pool from rabbits in our colony are present in such low amounts as to be chemically undetectable; a 2-fold, or even higher, increase in their relative concentration would probably also be undetectable. The IgA and IgG pools used in our study were obtained from the same colony of rabbits, and so the difference between the α - and γ -chain pools does not reflect differences between individual rabbits or rabbit groups. A similar quantitative difference between the heavy-chain subgroups of human IgA and IgG has been reported (Capra & Kehoe, 1975b).

The existence of variable-region structural correlates of the serologically detected a -locus allotypes has been confirmed in the present study, and it is argued that the uniform variable-region structure of all three allotypes and of α -negative molecules could only be maintained in a small number of germ-line genes. All of our data support the Translocation Hypothesis; the allotypic and isotypic variants of the variable region were found to be structurally identical in the α - and γ -chain pools. Further experimentation is required to explain the uneven distribution of variable-region subgroups throughout the heavy-chain classes.

We thank Professor R. R. Porter for his encouragement throughout this study and his advice, together with that of Dr. E. M. Press and Dr. A. F. Williams, in the preparation of this manuscript. Our thanks are also due to Mr. A. C. Willis for excellent technical assistance, to Ms. J. Addis for automated sequence analyses, and to Mr. T. Gascoyne for amino acid analyses. We are most grateful to Dr. G. R. Drapeau for the gift of 'V8-protease'.

References

- Ansari, A. A., Carta-Sorcini, M., Mage, R. G. & Appella, E. (1976) *J. Biol. Chem.* **251**, 6798–6806.
- Bennett, J. C. (1968) *Biochemistry* **7**, 3340–3344.
- Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029–3035.
- Capra, J. D. & Kehoe, J. M. (1975a) *Adv. Immunol.* **20**, 1–40.
- Capra, J. D. & Kehoe, J. M. (1975b) *J. Immunol.* **114**, 678–681.
- Capra, J. D., Wasserman, R. L. & Kehoe, J. M. (1973) *J. Exp. Med.* **138**, 410–427.
- Dixon, H. B. F. & Perham, R. N. (1968) *Biochem. J.* **109**, 312–314.
- Dreyer, W. J. & Bennett, J. C. (1965) *Proc. Natl. Acad. Sci. U.S.A.* **54**, 864–869.
- Gally, J. A. & Edelman, G. M. (1970) *Nature (London)* **227**, 341–348.
- Gray, W. R. (1967) *Methods Immunol.* **11**, 469–475.
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822.
- Hill, R. L., Lebovitz, H. E., Fellows, R. E., Jr. & Delaney, R. (1967) in *Nobel Symp.: Gamma Globulins, Structure and Control of Synthesis* (Killander, J., ed.), vol. 3, pp. 109–127, Almquist and Wiksell, Stockholm.
- Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3506–3509.
- Jaton, J.-C. (1974) *Biochem. J.* **143**, 723–732.
- Jaton, J.-C. (1975) *Biochem. J.* **147**, 235–247.
- Jaton, J.-C., Braun, D. G., Strosberg, A. D., Haber, E. & Morris, J. E. (1973) *J. Immunol.* **111**, 1838–1843.
- Johnstone, A. P. & Mole, L. E. (1975) *Nature (London)* **256**, 339–340.
- Johnstone, A. P. & Mole, L. E. (1977) *Biochem. J.* **167**, 245–253.
- Kim, B. S. & Dray, S. (1972) *Eur. J. Immunol.* **2**, 509–514.
- Köhler, H., Shimizu, A., Paul, C., Moore, V. & Putnam, F. W. (1970) *Nature (London)* **227**, 1318–1320.
- Kostka, V. & Carpenter, F. H. (1964) *J. Biol. Chem.* **239**, 1799–1803.
- Mole, L. E. (1975) *Biochem. J.* **151**, 351–359.
- Mole, L. E. (1976) *Biochem. Soc. Trans.* **4**, 33–34.
- Mole, L. E., Jackson, S. A., Porter, R. R. & Wilkinson, J. M. (1971) *Biochem. J.* **124**, 301–318.
- O'Donnell, I. J., Frangione, B. & Porter, R. R. (1970) *Biochem. J.* **116**, 261–268.
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593.
- Pan, S. C. & Dutcher, J. D. (1956) *Anal. Chem.* **28**, 836–838.
- Prahl, J. W., Tack, B. F. & Todd, C. W. (1973) *Biochemistry* **12**, 5181–5186.
- Pratt, D. M. & Mole, L. E. (1975) *Biochem. J.* **151**, 337–349.
- Smithies, O., Gibson, D., Fanning, E. M., Goodfleisch, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921.
- Summers, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624–628.
- van Hoergarden, M. & Strosberg, A. D. (1976) *FEBS Lett.* **66**, 35–38.
- Wilkinson, J. M. (1969) *Biochem. J.* **112**, 173–185.
- Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **133**, 369–370.