Amino Acid Sequence of the (λ) Light Chain of a Human Myeloma Immunoglobulin (IgG New)†

Betty L. Chen and Roberto J. Poljak*,‡

ABSTRACT: The amino acid sequence of the light (λ) chain Newm from immunoglobulin IgG New has been determined by chemical methods, by sequence homology with other human λ chains, and by information obtained from high-resolution X-ray diffraction studies. The sequence determination was carried out mainly on the tryptic peptides of the aminoethylated protein. The pattern of variation which is observed in the variable N-terminal half of the chain fits well with that observed in other human (λ) chains, with the exception of a deletion of seven amino acids at positions 54–60. The sequence of the constant C-terminal half of the molecule is in agreement with that of a typical λ chain. Positions 154 (serine) and 191 (lysine) correspond to the serological markers Kern⁻ and Oz⁺, respectively.

The Fab′ fragment of the human myeloma immunoglobulin IgG New[1] has been crystallized (Rossi *et al.* 1969) and its three-dimensional structure has been determined at 6-Å resolution (Poljak *et al.*, 1972), and more recently at 2.8-Å resolution (Poljak *et al.*, 1973). The study of the light (L) chain of IgG New reported here was undertaken to obtain the amino acid sequence information necessary for a complete interpretation of the 2.8-Å Fourier map of Fab′ New. The amino acid sequences of the light (L) chains of immunoglobulins are usually divided into two homology regions, V_L and C_L, consisting of approximately 110 amino acids each, which correspond to the variable N-terminal half and to the constant C-terminal half, respectively (see Dayhoff (1972) for a suitable compilation of these sequences). Human L chains show identity (or near identity) of sequence in their C_L regions and partial identity and strong homology in their V_L regions. Thus, in the study of λ chain Newm, sequence homologies could be used to align many of the peptides obtained by enzymatic cleavage and to assign to them, based on their amino acid compositions, an expected or probable amino acid sequence. Furthermore, since this study was undertaken in conjunction with the high-resolution X-ray diffraction analysis of Fab′ New, information obtained from the Fourier map was also used to help in the interpretation of the chemical sequence data.

Experimental Section

Materials. Human myeloma immunoglobulin New (IgG1, Gm (1 + 3− 4− 5−), λ) was purified from New serum kept frozen at −20°. Diethylaminoethyl cellulose (0.9 mequiv/g) and carboxymethyl cellulose (0.8 mequiv/g) were purchased from Serva. Sephadex G-100 and G-25 were obtained from Pharmacia. TPCK trypsin, chymotrypsin, carboxypeptidase A (treated with DFP), carboxypeptidase B, and pepsin were purchased from Worthington Biochemical Corp. Thermolysin was obtained from Calbiochem. Dithiothreitol, ethylenimine, iodoacetic acid, dansyl, and sequence reagents were obtained from Pierce Chemical Co. and used fresh without further purification. Polyamide thin-layer sheets were purchased from Gallard-Schlesinger. Dansyl amino acid standards were prepared as described by Gray (1967a).

Methods. Preparation of L Chain Newm. IgG New was purified by precipitation with sodium sulfate and chromatography on diethylaminoethyl cellulose and carboxymethyl cellulose as described before (Rossi and Nisonoff, 1968). Mildly reduced and alkylated L chain was prepared by reducing a 20-mg/ml solution of IgG New in 0.2 M Tris-HCl (pH 8.5)–0.005 M EDTA–0.02 M dithiothreitol for 2 hr at room temperature, followed by alkylation with iodoacetic acid added to the reaction mixture to a final concentration of 0.06 M. Alkylation was continued for 20 min in the dark, with active stirring, and was followed by dialysis (2–4 hr) against 1 M propionic acid, also in the dark. Preparative separation of heavy (H) and L chains was accomplished using 5 × 100 cm Sephadex G-100 columns equilibrated with 1 M propionic acid and fitted with upflow adaptors. The H and L chain fractions were pooled, lyophilized, and stored as a dried powder. Totally reduced and alkylated L chain was prepared by reduction in 8 M guanidine-HCl, 0.2 M Tris-HCl (pH 8.5), 0.005 M EDTA, and 0.1 M dithiothreitol for 2 hr at 40°. After alkylation with iodoacetic acid (added to a final concentration of 0.3 M) the sample was dialyzed against freshly prepared 8 M urea–1 M acetic acid buffer and filtered through a 5 × 100 cm Sephadex G-100 column equilibrated with the same buffer. The H and L chains thus obtained were dialyzed against water and stored as lyophilized powders.

Enzymatic Digestion of L Chain. The mildly reduced and alkylated L chain was performic acid oxidized or aminoethylated following the procedures of Hirs (1967) and of Raftery and Cole (1966), respectively. Aminoethylated or performic acid oxidized L chain was digested with TPCK trypsin or chymotrypsin. Digestions were carried out for 4 hr at room temperature in 0.1 M NH_4HCO_3 using an enzyme:L chain

† From the Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205. *Received October 19, 1973.* This work was supported by research Grants AI 08202 from the National Institutes of Health and E-638 from the American Cancer Society.

‡ Research Career Development awardee of the National Institutes of Health (AI-70091).

[1] Abbreviations used for immunoglobulins, their chains, and fragments are as recommended in *Bull. W. H. O. 30,* 447 (1964). λ_{Newm} designates the light polypeptide chain obtained from IgG New, λ_{Newm} rather than λ_{New} is used to avoid confusion with another human λ chain of the same designation that has been sequenced before (Langer *et al.*, 1968). Other abbreviations used are: dansyl, 1-dimethylaminonaphthalene-5-sulfonyl; DFP, diisopropyl fluorophosphate; EDTA, disodium ethylenedinitrilotetraacetate; PITC, phenyl isothiocyanate; PCA, pyrrolidonecarboxylic acid; TPCK, 1-tosylamido-2-phenylethyl chloromethyl ketone.

TABLE I: Composition of λ Chain Newm from Amino Acid Analysis and from Sequence Analysis.[a]

Amino Acid	Amino Acid[b] Analysis	Sequence
Lysine	12.70	13
Histidine	4.10	4
Arginine	5.05	5
Half-cystine	5.02	5
Aspartic acid	12.00	12
Threonine	19.26[c]	19
Serine	30.34[c]	30
Glutamic acid	21.08	21
Proline	14.32	14
Glycine	14.13	14
Alanine	19.05	19
Valine	15.85	16
Methionine	0.00	0
Isoleucine	4.57	5
Leucine	15.08	15
Tyrosine	7.94[c]	8
Phenylalanine	4.73	5
Tryptophan	N.d.[d]	3
Total		208

[a] Values listed are moles per mole of L chain. [b] Average of ten analyses performed on performic acid oxidized, carboxymethylated and aminoethylated, L chain samples. [c] Corrected for partial destruction during hydrolysis using an experimental, time dependent factor. [d] Not determined.

ratio of 1:100. A slight precipitate that developed during the course of trypsin digestion was removed by centrifugation. Digestion of purified peptides with chymotrypsin, trypsin, or pepsin was performed as proposed by Smyth (1967). A similar procedure was used for thermolysin digestions which were conducted at 45°. Digestions with carboxypeptidases A and B were performed as described by Ambler (1967).

Fractionation and Purification of Peptides. Peptides obtained by digestion of L chain with trypsin or chymotrypsin were separated by gel filtration in 2.5 × 100 cm columns of Sephadex G-25 (fine) equilibrated with 0.1 M ammonia. The effluent of the columns was monitored by absorption readings at 280 and 232 nm. By this procedure the tryptic and chymotryptic digests were each divided into four fractions. The fractions were lyophilized, redissolved in pH 6.5 buffer, and submitted to high-voltage electrophoresis at pH 6.5 using gradients of 50 V/cm for periods of time ranging from 60 to 400 min in varsol cooled tanks. Cysteic acid, glutamic acid, aspartic acid, threonine, leucine, histidine, arginine, and lysine were used as electrophoretic markers and run side by side with the peptides under investigation. After electrophoresis, guide strips were stained with ninhydrin (0.2% in acetone), with a chlorine stain (Reindel and Hoppe, 1954), and with stains specific for arginine and tryptophan (Bennet, 1967). Subsequent purification steps included paper chromatography and high-voltage electrophoresis at different pH values. The papers used in chromatography were previously washed by descending chromatography with distilled water. The chromatographic solvent consisted of 1-butanol–acetic acid–pyridine–water (25:3:10:12). Further peptide purification was achieved by high-voltage electrophoresis using a pH 2.0 buffer (formic acid–acetic acid–water = 25:87:888), a pH 4.7 buffer (acetic acid–pyridine–water = 1:1:78), and a pH 3.5 buffer (acetic acid–pyridine–water = 10:1:189). When a two-dimensional or a "three-dimensional" separation procedure was required to purify a given peptide, its location on paper was established using a weak ninhydrin stain (0.02% ninhydrin in acetone). As soon as the peptide became visible the area containing it was cut and washed with a 1:1 mixture of ethanol–acetone. Peptides were eluted from paper (using 6 N HCl) into 50- or 100-μl glass capillaries which were sealed and hydrolyzed at 110° (18–24 hr) for amino acid analysis. Alternatively, peptides were eluted with 0.1–0.3 ml of a 1:1 mixture of water–pyridine and reserved for sequence determination.

Amino Acid Analyses. Qualitative amino acid analyses were performed by high-voltage electrophoresis at 60 V/cm for 90 min using Whatman 3MM paper and the pH 2.0 buffer described above. The paper was stained with ninhydrin (0.2% in acetone)–acetic acid–trimethylpyridine (50:15:2). Quantitative amino acid analyses were performed by the method of Moore *et al.* (1968) using a Beckman Spinco 120C amino acid analyzer.

Sequencing Techniques. Sequential degradation of purified peptides was performed using the Edman reaction (Edman, 1956). After each degradation step, aliquots of peptides were removed to determine the newly exposed N-terminal residue, as proposed by Hartley (1970) and Gray (1967b). The dansyl derivatives of amino acids were identified by chromatography using the procedure of Woods and Wang (1967) on 4.5 × 7.5 cm polyamide sheets. Chromatography was performed using 1.5% formic acid in the first dimension, benzene–glacial acetic acid (9:1) in the second dimension, and ethyl acetate–methanol–acetic acid (20:1:1) in the third dimension.

Results

Amino Acid Composition of the L Chain. The amino acid composition of the aminoethylated L chain from IgG New is given in Table I. The listed value of five half-cystine residues is compatible with the well-established model of immunoglobulin L chains which contain two intrachain disulfide bonds and one interchain disulfide link. As shown in Table I, no methionine was detected by amino acid analysis. Determination of the N-terminal amino acid using the dansyl technique was negative, indicating the presence of a blocked N-terminus. A blocked N-terminus and the absence of methionine residues are frequently observed in human λ chains. These two features limited to a large extent the technical approach to the determination of the amino acid sequence which was obtained mainly from the tryptic peptides of the aminoethylated protein.

Characterization and Sequence of the Tryptic Peptides. The amino acid compositions of the tryptic peptides that were purified and analyzed are given in Table II (the numbers used to designate tryptic peptides are those of Figure 1). The sum of the integral number of amino acids derived for the tryptic peptides is close to the experimental values determined by amino acid analysis of the whole protein (Table I). Tables III–VI summarize the sequence information obtained from the tryptic peptides. Additional data for some of the peptides are given in the following paragraphs.

T1. The presence of a blocked N-terminus suggested that this (ninhydrin negative) peptide contains the N-terminal sequence of L chain Newm. After digestion of T1 with thermolysin, three major subpeptides T1-th2, T1-th4, and T1-th10 were separated by high-voltage electrophoresis at pH 2.0, followed by descending paper chromatography. The following sequences were obtained by use of the dansyl-Edman procedure.

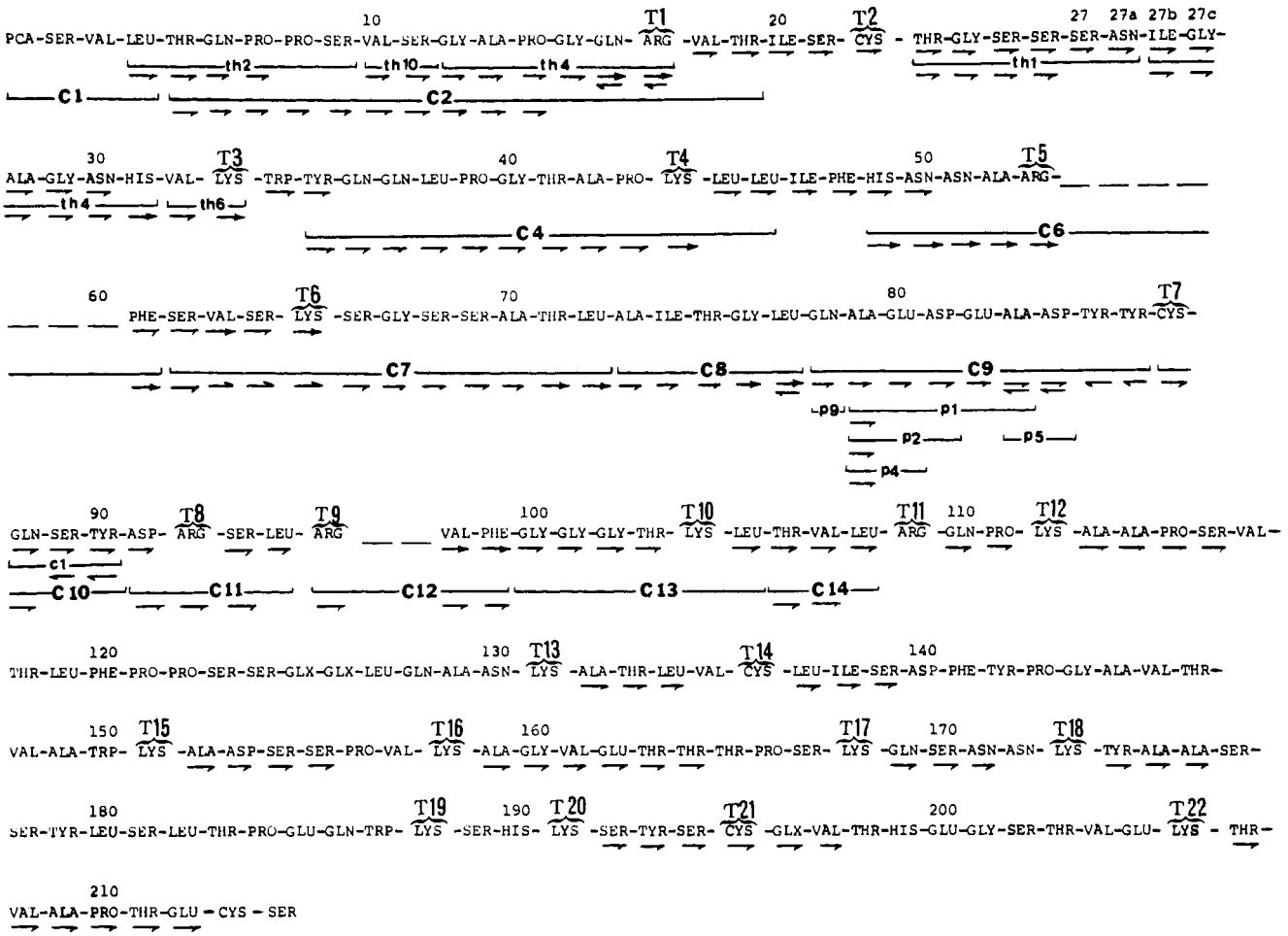


FIGURE 1: Summary of sequence information obtained for the L (λ) chain Newm. Tryptic and chymotryptic peptides are indicated by T and C, respectively. Subpeptides obtained by digestion of tryptic peptides with thermolysin (th) and chymotrypsin (c) are indicated. Pepsin subpeptides of chymotryptic peptide C9 are designated by p. Other symbols used are: →, ⇀, and ⇁ for residues identified by the dansyl Edman, the subtractive Edman, or both techniques, respectively: ← for residues identified by digestion with carboxypeptidases (A or B); all other residues are placed by sequence homology and by information obtained from a Fourier map of the electron density at 2.8-Å resolution (Poljak *et al.*, 1973). Numbers 27a, 27b, and 27c, and gaps (57–63 and 96–97) are introduced to maximize homology with other human λ chains.

T1-th2: Leu-Thr-Gln (Pro,Ser)

T1-th4: Gly-Ala-Pro-Gly-Gln-Arg,
C-terminal sequence determined using carboxypeptidases A and B (4 hr digest)

T1-th10: Val-Ser

T3. Three major subpeptides obtained by thermolysin digestion of T3 were purified by high-voltage electrophoresis at pH 2.0 followed by chromatography. The following sequences were obtained.

T3-th1: Thr-Gly-Ser-Ser-(Ser,Asn);
electrophoretic mobility at pH 6.5: neutral

T3-th4: Ile-Gly-Ala-Gly-Asn-His;
electrophoretic mobility at pH 6.5; basic

T3-th6: Val-Lys

T4. The determination of the N-terminal amino acid by the dansyl technique showed the HCl hydrolysis products usually associated with tryptophan. After one cycle of Edman degradation the ultraviolet absorption spectra of T4 showed a shift which was interpreted as due to the loss of a tryptophan residue. Also, since T4 contains tryptophan and all the other amino acids present in T4 can be assigned definite positions in its sequence (by overlap with the chymotryptic peptide C4; see Figure 1), we concluded that tryptophan is the N-terminal amino acid. Because of side products observed after two Edman degradation steps of T4, the sequence in this region of the structure was continued on the chymotryptic overlap peptide C4.

T6. Since no Val residue has been observed in other human λ chain sequences at position 63, this residue was also determined by the subtractive Edman technique.

T7. This is a large, acidic peptide. A similar, slightly larger peptide containing cysteic acid and arginine was obtained from tryptic digests of the performic acid oxidized L chain. As these peptides were fairly insoluble and difficult to handle, sequence determination in this region of the molecule was performed on the chymotryptic peptides.

T8. After chymotrypsin digestion of T8 a (neutral) tripeptide (Gln,Ser)Tyr was isolated by electrophoresis at pH 6.5. After treatment of this tripeptide with carboxypeptidase A (4 hr), tyrosine and trace amounts of serine were detected by qualitative amino acid analysis (paper electrophoresis at pH 2.0) in agreement with the sequence

Gln-Ser-Tyr

Characterization of Chymotryptic Peptides of the L Chain. The amino acid composition of some of the purified peptides is given in Table VII. Peptides that were considered important for the determination of the sequence were characterized as summarized in Tables VIII and IX. Additional information for some of the peptides is given in the following paragraphs.

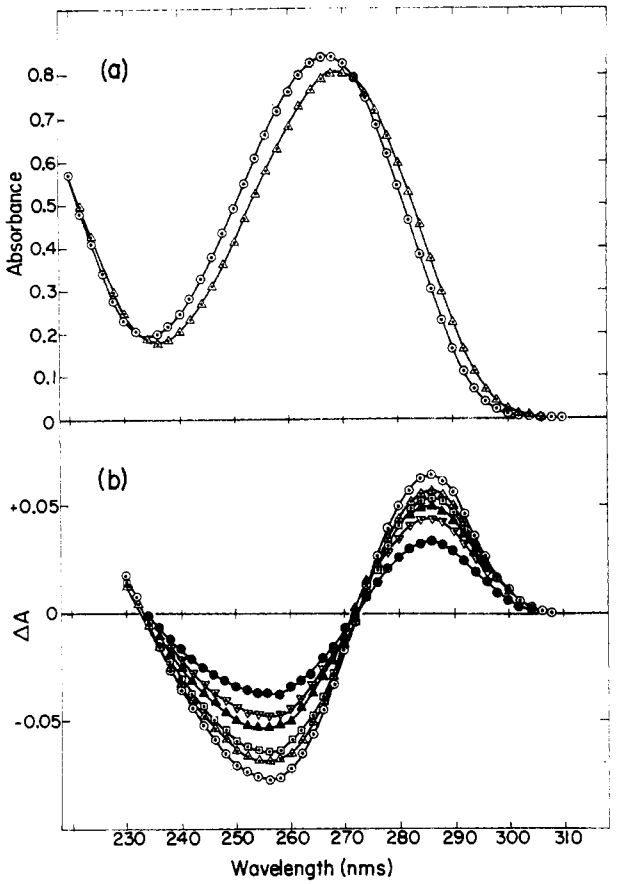
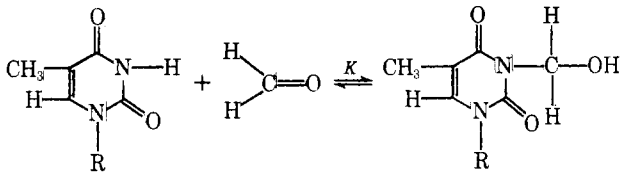


FIGURE 1: (a) Spectrum of thymidine: (○) control; (△) with 3.2 *M* formaldehyde. (b) Difference spectra of thymidine in the presence of various concentrations of added formaldehyde: (●) 0.3 *M*; (▽) 0.5 *M*; (▲) 0.7 *M*; (□) 1.1 *M*; (△) 1.3 *M*; (○) 2.1 *M* formaldehyde.

exocyclic amino group, but have an endocyclic "imino" group, can be written as follows (using thymidine as an example):



where the endocyclic N-3 proton is replaced by a hydroxymethyl or methylol group. The site of reaction is demonstrated by "formol titration" [in which formaldehyde is found to raise the apparent p*K* for the dissociation of this endocyclic proton (Lewin, 1964)], by analogy with the cyanoethylation reaction (Eyring and Ofengand, 1967), and by the observation (present study) that neither N^3-methyluridine (at neutral pH) nor 5′-TMP (at pH 12) shows any significant absorbance change on being incubated with formaldehyde. That the adduct is indeed a methylol or hydroxymethylol group is suggested from the arguments of the previous paper for the reaction with amino groups, by the similarity between the ultraviolet spectra of the reaction products and the corresponding methylated compounds, and by the chemical "unreasonableness" of the alternative cationic Schiff base.

(2) Equilibrium Constants and Spectral Parameters. Figure 1a shows the spectra for thymidine alone and in the presence of a high concentration (3.2 *M*) of formaldehyde; the spectral change on reaction with HCHO is quite small, but nevertheless significant. Difference spectra for thymidine at a number of intermediate formaldehyde concentrations (Figure 1b) show a peak at 286 nm, a valley at 256 nm, and a reasonably good isosbestic wavelength at 272–273 nm. As discussed in the preceding paper, the usual method of determining the equilibrium constant from such data is to measure, as a function of formaldehyde concentration, the absorbance changes at one particular wavelength (usually either at the peak or the valley of the difference spectra), and then to plot these changes according to standard mass action relations. However, for these small changes in absorbance such plots were found to be very sensitive to experimental error, and therefore not very effective in proving that there are indeed only two absorbing compounds in the system, as supposed in the above reaction scheme. To bypass these difficulties, a method of spectral analysis was devised in which a multi-wavelength spectrum S(F) of the nucleotide or nucleoside in the presence of a certain formaldehyde concentration, F, is expressed as a linear combination of two "base" spectra; S(O), the spectrum of unreacted starting compound, and S(H), the spectrum of the compound in the presence of a fixed high formaldehyde concentration, H. Algebraically

$$\mathbf{S}(F) = \alpha\mathbf{S}(O) + \beta\mathbf{S}(H)$$

where the coefficients α and β are functions of the formaldehyde concentration. It is shown in detail in the appendix that a plot of the ratio α/β vs. the reciprocal of the formaldehyde concentration should be linear, with equilibrium constants determined from the slope and intercept according to eq App-5. This method of analysis should be applicable to any two-component system in which the product spectrum is unobtainable, either by isolation or by saturation, and has the following advantages: the data from 25–30 wavelengths over the entire spectrum are used, with a considerable increase in the accuracy and reliability of the estimated equilibrium constants; the spectrum of the pure addition product can be obtained; and a consistency relation in the analysis indicates whether the experimental system does indeed conform to the two-component reaction scheme.

Typical data from such analyses are shown for 5′-TMP, 5′-UMP, and for poly(U) in Figure 2a–c, respectively; all plots are seen to be linear as required by eq App-5. Equilibrium constants and spectral parameters are collected in Table I for a variety of compounds with reactive endocyclic nitrogens, and lead to the following general observations. (i) Except for the last two entries in Table I, the estimated equilibrium constants are quite low (2–3 M^{-1}), and adduct spectra are generally shifted 2–3 nm toward longer wavelengths, accompanied by slight intensity decreases. The equilibrium constants agree within experimental error with those reported for uridine (Eyring and Ofengand, 1967) and for 5′-UMP (Aylward, 1966). (ii) The last two entries in Table I, dimethyladenine and purine, where the reaction presumably takes place at the N-9 position, appear to fall into a separate class characterized by considerably higher equilibrium constants ($\geq$10 M^{-1}), in reasonable agreement with values determined (by titration) for adenine and purine (Lewin and Barnes, 1966). (iii) Based on comparisons of nucleosides and nucleotides, there appears to be no significant effect of the 5′-phosphate group on either the equilibrium constant or the spectral changes accompanying adduct formation. (iv) Uracil and thymine derivatives show virtually identical equilibrium and spectral parameters. (v) Increased ionic strength (up to 1 *M* NaCl) has no effect on

FOOTNOTES TO TABLE II

[a] Values listed as residues/mol. Numbers in parentheses are the assumed integral values. [b] Experimental values not corrected for destruction. [c] Tryptophan was determined by specific stain (see Methods) and its value assumed to be 1. [d] Analyzed qualitatively (see Methods), values are estimated. [e] Estimated from amount of material digested. [f] Carboxymethylcysteine.

TABLE III: Characterization of Tryptic Peptides from the Sephadex G-25 Fraction I.[a]

T3	Sequence: Thy-Gly-Ser-Ser-Ser-Asn-Ile-Gly-Ala-Gly-Asn-(His,Val),Lys Electrophoretic mobility, pH 6.5: basic
T7	Purified by high-voltage electrophoresis at pH 4.7 and 6.5 barely soluble in electrophoresis buffer. See text.

[a] Symbols used in this and subsequent tables are: (——→) dansyl-Edman determination; (——▸) subtractive Edman step; (←——) carboxypeptidase digestion.

TABLE IV: Characterization of Tryptic Peptides From the Sephadex G-25 Fraction II.

T1	Ninhydrin negative, detected by chlorine and arginine stains Blocked N-terminus. See text.
T2	Sequence: Val-Thr-Ile-Ser-Cys (Cys is aminoethylcysteine)
T5	Sequence: Leu-Leu-Ile-Phe-His-Asn-(Asn,Ala),Arg
T13	Sequence: Ala-Ala-Pro-Ser-(Val,Thr,Leu,Phe,Pro,Pro,Ser,Ser,Glx,Glx,Leu,Glx,Ala,Asn),Lys
T15	Contains Trp. Sequence: Leu-Ile-Ser-(Asp,Phe,Tyr,Pro,Gly,Ala,Val,Thr,Val,Ala,Trp),Lys Electrophoretic mobility, pH 6.5: neutral
T16	Sequence: Ala-Asp-Ser-Ser-(Pro,Val),Lys Electrophoretic mobility, pH 6.5: neutral
T17	Sequence: Ala-Gly-Val-Glu-Thr-Thr-(Thr,Pro,Ser),Lys Electrophoretic mobility, pH 6.5: neutral
T19	Sequence: Tyr-Ala-Ala-(Ser,Ser,Tyr,Leu,Ser,Leu,Thr,Pro,Glx,Glx,Trp),Lys
T22	Sequence: Glx-Val-(Thr,His,Glx,Gly,Ser,Thr,Val,Glx),Lys
T23	Sequence: Thr-Val-Ala-Pro-Thr-Glx-(Cys,Ser) (Cys is carboxymethylcysteine)

The designation of the chymotryptic peptides is that shown in Figure 1.

C6. This peptide was obtained in high yields from digests of the aminoethylated and the performic acid oxidized L chain. Because of its key location in the structure (see Discussion) it was sequenced by the dansyl-Edman and the subtractive Edman techniques.

C9. Digestion of C9 with pepsin gave rise to several peptides, some of which were purified as very acidic bands by electrophoresis at pH 6.5. They were characterized as follows (listed in order of decreasing electrophoretic mobility)

C9-p1: Ala-(Glu,Asp,Glu,Ala)
C9-p2: Ala-(Glu,Asp)
C9-p4: Ala-(Glu)
C9-p5: (Ala,Asp)
C9-p9 Gln-(electrophoretic mobility at pH 6.5; neutral)

The electrophoretic mobility of the peptic fragments of C9 allows an assignment of amide positions in this region of the sequence as indicated in Figure 1.

Discussion

High-resolution Fourier maps of protein structures obtained by X-ray crystallographic analyses have been successfully interpreted in several laboratories by a correlation of electron density features with the primary structure of the protein under study. Many regions of these maps have been completely analyzed only when sequence information became available. Although the work presented in this paper was not intended to be a complete amino acid sequence determination by chemical techniques, it provided sequence information which was necessary to interpret the three-dimensional structure of the L chain obtained in a 2.8-Å Fourier map of Fab′ New (Poljak *et al.*, 1973). Some aspects of the amino

TABLE V: Characterization of Tryptic Peptides from Sephadex G-25 Fraction III.

T4	Sequence: Trp-Tyr-(Gln,Gln,Leu,Pro,Gly,Thr,Ala,Pro),Lys Electrophoretic mobility, pH 6.5: basic
T6	Sequence: Phe-Ser-Val-Ser-Lys
T8	Sequence: Glx-Ser-Tyr-Asx-Arg
T10	Sequence: Val-Phe-Gly-Gly-Gly-Thr-Lys
T11	Sequence: Leu-Thr-Val-Leu-Arg
T14	Sequence: Ala-Thr-Leu-(Val),Cys (Cys is aminoethylcysteine)
T18	Sequence: Glx-Ser-Asx-(Asx),Lys

TABLE VI: Characterization of Tryptic Peptides from Sephadex G-25 Fraction IV.

T9	Sequence: Ser-Leu-Arg
T12	Sequence: Gln-Pro-Lys Electrophoretic mobility, pH 6.5: basic
T20	Shows yellow color immediately after ninhydrin stain Most probable sequence: Ser-His-Lys
T21	Sequence: Ser-Tyr-Ser-Cys (Cys is aminoethylcysteine)

TABLE VII: Amino Acid Composition of Chymotryptic Peptides of L Chain Newm.[a]

Amino Acid	C1	C2	C4	C6	C7	C8	C9	C10	C11	C12	C13	C14
Lysine			0.96 (1)		0.86 (1)						1.00 (1)	
Histidine				1.01 (1)								
Arginine		0.91 (1)		0.99 (1)					0.88 (1)	0.87 (1)		
AE-cysteine								1.00 (1)				
Aspartic acid				2.21 (2)			2.07 (2)		0.94 (1)			
Threonine		1.65 (2)	0.84 (1)		0.99 (1)	0.85 (1)					1.07 (1)	1.00 (1)
Serine	0.99 (1)	2.19 (2)			4.75 (5)			0.86 (1)	1.02 (1)			
Glutamic acid	1.08 (1)	2.00 (2)	1.65 (2)				2.96 (3)	1.13 (1)				
Proline		3.17 (3)	1.87 (2)									
Glycine		1.98 (2)	1.05 (1)		1.00 (1)	1.00 (1)					2.63 (3)	
Alanine		1.06 (1)	1.10 (1)	1.01 (1)	1.00 (1)	0.85 (1)	1.95 (2)					
Valine	0.88 (1)	1.88 (2)			0.97 (1)					1.15 (1)		0.92 (1)
Isoleucine						1.02 (1)						
Leucine	0.94 (1)		3.13 (3)		1.01 (1)	0.97 (1)			1.02 (1)		1.04 (1)	1.07 (1)
Tyrosine[b]			0.73 (1)				1.35 (2)	0.93 (1)				
Phenylalanine				1.06 (1)						0.97 (1)		
Tryptophan												
Total residues	4	15	12	6	11	5	9	4	4	3	6	3
% yield[c]	30	21	6	10	36	23	12	11	38	28	13	36
Residue no.	1–4	5–19	35–46	49–61	62–72	73–77	78–86	87–90	91–94	95–99	100–105	106–109

[a] Values listed as residues/mol. Numbers in parentheses are the assumed integral values. [b] Experimental values not corrected for destruction. [c] Estimated from amount of material digested.

TABLE VIII: Characterization of Chymotryptic Peptides from Sephadex G-25 Fractions I and II.

C2	Sequence: Thr-Gln-Pro-Pro-Ser-Val-Ser-Gly-Ala-Pro-(Gly,Gln,Gln,Arg,Val,Thr) Electrophoretic mobility, pH 6.5: basic
C4	Sequence: Tyr-Gln-Gln-Leu-Pro-Gly-Thr-Ala-Pro-Lys-(Leu,Leu) Electrophoretic mobility, pH 6.5: basic
C7	Sequence: Ser-Val-Ser-Lys-Ser-Gly-Ser-Ser-Ala-Thr-Leu
C9	Sequence: Glx-Ala-Glx-Asx-Glx-Ala-Asp-Tyr-Tyr Carboxypeptidase A, 1 hr: Tyr, 1.54, Asp, 0.44; Ala, 0.18 18 hr: Tyr, 2.02; Asp, 0.60; Ala, 0.25 Electrophoretic mobility, pH 6.5: very acidic

acid sequence presented here (Figure 1) which are directly related to this goal as well as other relevant features will be briefly discussed in the following paragraphs.

The N-terminal sequence of L chain Newm is assumed to be PCA-Ser-Val-, in agreement with the amino acid composition data, with the relatively constant, homologous N-terminal sequence of other human λ chains, and with the high-resolution Fourier maps of Fab' New (Poljak *et al.*, 1973) in which an N-terminal PCA ring appears identifiable. Residues 27a, 27b, and 27c (Figure 1), which are deleted in many human chain sequences, and the adjacent "hypervariable positions" (Wu and Kabat, 1970) can be satisfactorily sequenced since they are included in one tryptic peptide. Within this peptide, position 31, which is most frequently occupied by tyrosine in human, pig, and mice λ chains, is replaced by histidine. Tyrosine-31 has been tagged by affinity labeling techniques in several laboratories (see Singer *et al.*, 1971) and appears contiguous to the active site of antibody molecules. In agreement with this conclusion, histidine-31 is exposed to the solvent in the three-dimensional structure of Fab' New. The tryptophan residue at the beginning of the tryptic peptide T4 (positions 34–44) can be easily recognized in the Fourier map of Fab' New. A tryptophan residue at position 34 is a constant feature of immunoglobulin L chain sequences and a constant (or nearly constant) feature of the homology regions of H and L chains. These tryptophan residues occur 14–16 residues after the first cysteine residue of the intrachain disulfide bond and are conspicuously contiguous to that bond in the three-dimensional structure of all the homology subunits of Fab' New.

When the amino acid sequence of L chain Newm is compared to that of other human chains a gap or deletion of seven amino acids is found at positions 54–60. This deletion has not been found in other λ chains which have been studied. Three lines of evidence were obtained in support of this finding. First, a tryptic peptide which is usually found in the corresponding region of other human λ chains was not found in digests of L chain Newm despite a prolonged search. The adjoining tryptic peptides T5 (positions 45–53) and T6 (positions 61–65), however, were obtained in high yields, so that it would be expected that the missing peptide (positions 54–60) should have been released in equally high yields by trypsin action. The homologous peptide from a human myeloma λ chain P.W., also under study in this laboratory, was easily located and characterized using the same techniques that were applied with λ chain Newm. The second experimental evidence indicating the deletion at positions 54–60 is the sequence of the chymotryptic peptide C6 (positions 49–61, Figure 1), which spans the length of the deletion and provides an overlap for the sequences of the tryptic peptides T5 and T6. Although C6 only provides a one residue overlap with T6, its sequence is in agreement with the reported deletion. It may be argued here that some peculiar sequence between positions 54 and 60, which would give rise, after proteolytic digestion, to single amino acids or short peptides, could have been misplaced or missed altogether and that this experimental difficulty has led to a wrong conclusion. The sequences of T5, T6, and C6, however, place a limitation on the nature of the possibly missing amino acids which would have to be, beginning at position 54, Phe-X-Arg, or Phe-X-Lys (where X = 0,1,2...*n* other amino acids) or a multiple repeat number of these sequences. This possibility appears highly unlikely since the amino acid composition of the L chain is accounted for by the sum of the amino acid residues of the purified tryptic peptides (Table I). In addition, these postulated sequences should have been detected in both the trypsin and chymotrypsin digests. The third experimental evidence indicating the deletion at positions 54–60 comes from the study of the 2.8-Å Fourier map of Fab' New (Poljak *et al.*, 1973). Due to the presence of residues with aromatic side chains (Phe-48, His-49, and Phe-61) and other distinctive residues (Lys, Leu, Ile, Arg, etc.), the polypeptide chain can be traced unequivocally in this region of the map. Only one sequence of covalently linked Arg-Phe residues can be placed in this region, and the physical continuity of the sequences corresponding to the tryptic peptides T5 and T6 is clearly established. The deletion of amino acid residues 54–61 modifies the folding of the V_L subunit, making it closer to that of the C_L and C_H1 subunits, as discussed before (Poljak *et al.*, 1973). It remains to be seen whether this deletion is peculiar to the structure of L chain Newm or whether it is a more general feature which will be found to occur in other immunoglobulin L chains.

TABLE IX: Characterization of Chymotryptic Peptides from Sephadex G-25 Fractions III and IV.

C1	Blocked N-terminal. Detected by chlorine stain Electrophoresis at pH 6.5: acidic
C6	Sequence: His-Asn-Asn-Ala-Arg-Phe Electrophoretic mobility, pH 6.5: basic
C8	Sequence: Ala-Ile-Thr-Gly-Leu Carboxypeptidase A (4-hr digest)
C10	Sequence: Cys-Gln-(Ser),Tyr Electrophoretic mobility, pH 6.5: basic
C11	Sequence: Asp-Arg-Ser-Leu Electrophoretic mobility, pH 6.5: neutral
C12	Sequence: Arg-Val-Phe
C14	Sequence: Thr-Val-Leu

The third region of hypervariable sequences of L chains occurs next to positions 96 and 97, where a gap or deletion is introduced (Figure 1) to align the L Newm sequence with that of other λ chains. Gaps or deletions of one or two amino acids have been reported before (Dayhoff, 1972) in human λ chains. However, it should be considered that the sequence of this region of the L chain Newm presents some technical difficulties since it is cleaved into small tryptic and chymotryptic peptides of limited overlap, a situation which could conceivably lead to the presence of one or two undetected residues. Here again we use the crystallographic evidence to support the sequence of this region as given in Figure 1.

Sequence determination in the C region of L chain Newm was limited to the objective of verifying the presence of the tryptic peptides that are expected on the basis of the known (constant) sequence of the C_L region of λ chains. Amino acid analysis and partial characterization of the tryptic peptides (Figure 2) show that C_L Newm conforms to this sequence. Positions 154 and 191 are occupied by serine and lysine, which correspond to the serological markers $Kern^-$ (Hess and Hilschmann, 1970) and Oz^+ (Appella and Ein, 1967), respectively. Since the C_L sequence agrees well with the electron density map of Fab' New it was accepted as a firmly established sequence for L chain Newm.

In general, the amino acid sequence of the L chain Newm conforms to the pattern of variation observed in other human λ chains that have been sequenced (Dayhoff, 1972). Residues which are known to be constant in both the V_L and C_L regions of human λ chains have also been found to occur in λ Newm, with the exception of those involved in the structural deletion (positions 54–61) discussed above. Assignment of L chain

Newm to one of the proposed subgroups of λ chains (Smith *et al.* 1971; Baczko *et al.*, 1970) appears difficult and will be postponed for a future report. The correlation of the amino acid sequence with the three-dimensional structure of the molecule has been discussed before (Poljak *et al.*, 1973) and will be reexamined in the near future using information derived from a 2-Å Fourier map of Fab' New which is currently under study in this laboratory.

Acknowledgment

We thank Dr. Samuel H. Boyer IV, who kindly provided the quantitative amino acid analyses reported in this work.

References

Ambler, R. P. (1967), *Methods Enzymol. 11*, 155.
Appella, E., and Ein, D. (1967), *Proc. Nat. Acad. Sci. U. S. 57*, 1449.
Baczko, K., Braun, D. G., Hess, M., and Hilschmann, N. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351*, 763.
Bennet, J. C. (1967), *Methods Enzymol. 11*, 330.
Dayhoff, M. O. (1972), Atlas of Protein Sequence and Structure, Vol. 5, Silver Spring, Md., National Biomedical Research Foundation.
Edman, P. (1956), *Acta Chem. Scand. 10*, 761.
Gray, W. R. (1967a), *Methods Enzymol. 11*, 139.
Gray, W. R. (1967b), *Methods Enzymol. 11*, 469.
Hartley, B. S. (1970), *Biochem. J. 119*, 805.
Hess, M., and Hilschmann, N. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351*, 67.
Hirs, C. H. W. (1967), *Methods Enzymol. 11*, 197.
Langer, B., Steinmetz-Kayne, M., and Hilschmann, N. (1968), *Hoppe-Seyler's Z. Physiol. Chem. 349*, 945.
Moore, S., Spackman, D. H., and Stein, W. H. (1968), *Anal. Chem. 30*, 1185.
Poljak, R. J., Amzel, L. M., Avey, H. P., Becka, L. N., and Nisonoff, A. (1972), *Nature (London), New Biol. 235*, 137.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P., and Saul, F. (1973), *Proc. Nat. Acad. Sci. U. S. 70*, 3305.
Raftery, M. A., and Cole, R. D. (1966), *J. Biol. Chem. 241*, 3457.
Reindel, T., and Hoppe, W. (1954), *Chem. Ber. 87*, 1103.
Rossi, G., Choi, T. K., and Nisonoff, A. (1969), *Nature (London) 223*, 837.
Rossi, G., and Nisonoff, A. (1968), *Biochem. Biophys. Res. Commun. 31*, 914.
Singer, S. J., Martin, N., and Thorpe, N. O. (1971), *Ann. N. Y. Acad. Sci. 190*, 342.
Smith, G. P., Hood, L., and Fitch, W. M. (1971), *Annu. Rev. Biochem. 40*, 969.
Smyth, D. G. (1967), *Methods Enzymol. 11*, 214.
Woods, K. R., and Wang, K. T. (1967), *Biochim. Biophys. Acta 133*, 369.
Wu, T. T., and Kabat, E. A. (1970), *J. Exp. Med. 132*, 211.

Catalysis by Chymotrypsinogen. Demonstration of an Acyl-Zymogen Intermediate†

Arieh Gertler,‡ Kenneth A. Walsh, and Hans Neurath*

ABSTRACT: Bovine chymotrypsinogen reacts stoichiometrically with diisopropyl phosphorofluoridate (DFP) and catalyzes the hydrolysis of the ester *p*-nitrophenyl *p*-guanidinobenzoate (NPGB). These reactions are mutually exclusive, indicating that the same residue on the active site (Ser-195) is involved in both cases. Although the second-order rate of ester hydrolysis by chymotrypsinogen is 10^6–10^7 times slower than by chymotrypsin, the reaction proceeds in both cases *via* the formation of acyl intermediates. The deacylation rate of the isolated acyl-enzyme is only 70 times slower than that of acyl-enzyme and is dependent upon the ionization of a single group, presumably His 57. The hydrolysis of NPGB by the homologous zymogen–enzyme pair, trypsinogen–trypsin, is competitively inhibited by *p*-aminobenzamidine. The apparent inhibition constant K_i is almost four orders of magnitude higher for the zymogen than for the enzyme. These data suggest that the inferior catalytic properties of the zymogens of the pancreatic serine proteases are primarily due to an undeveloped binding site and only secondarily to a less efficient catalytic apparatus. Circular dichroic spectra of acyl-chymotrypsinogen at pH 4.0 indicate that the changes in ellipticity in the 220–250-nm range induced by acylation resemble those induced in the enzyme. In the range of 260–290 nm, however, the spectrum of the acyl-zymogen is significantly different, indicating a perturbed environment of the *p*-guanidinobenzoyl group and suggesting a different and possibly less efficient mode of binding. These spectral changes are completely reversible upon deacylation.

The zymogens of several proteolytic enzymes react slowly with site-specific reagents and catalyze certain zymogen–enzyme transformations. Foltmann (1966) first showed that the activation of prorennin at low pH was catalyzed by the zymogen itself. Lacko and Neurath (1970), Reeck and Neurath (1972), Uren *et al.* (1972), and Behnke and Vallee (1972) demonstrated an intrinsic activity of procarboxypeptidases A and B toward ester and peptide substrates of the respective

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received October 26, 1973.* This work has been supported by research grants from the National Institutes of Health (GM 15731). A preliminary account has been published (Gertler, 1973).

‡ On leave of absence from Department of Biochemistry, Faculty of Agriculture, Hebrew University of Jerusalem, Rehovot, Israel.