chemistry 16 (following paper in this issue).
Underfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W., and Weigele, M. (1972), *Science 178*, 871-872.
Weinheimer, P. F., Mestecky, J., and Acton, R. T. (1971), *J. Immunol. 107*, 1211.
Woods, K. R., and Wang, K. T. (1967), *Biochim. Biophys. Acta 133*, 369.

Primary Structure of Human J Chain: Alignment of Peptides from Chemical and Enzymatic Hydrolyses†

John Edwin Mole,* Ajit Singh Bhown, and J. Claude Bennett

ABSTRACT: The primary structure of the J chain from a human Waldenströms IgM protein has been determined using a combination of automated and conventional Edman degradative procedures. Eighty-five percent of the sequence was established with peptides isolated from tryptic digests of carboxyamidomethylated and citraconylated J chain, many of which were sequenced completely. Alignment of the tryptic fragments was achieved with peptides generated by chymotrypsin and limited acid hydrolyses. The J chain consists of 129 amino acids and a single oligosaccharide structure linked to asparagine at position 43 of the sequence. The molecular weight, including 7.5% carbohydrate by weight, is 16 422. The location and arrangement of three half-cystines could be deduced from previous studies, whereas the pairing of the remaining five disulfide bonds still needs to be clarified.

Immunoglobulin M has been shown to consist of five 7S subunits which are joined by disulfide bonds into a complex macromolecule. Similarly, IgA may be organized into polymeric patterns, although a monomeric form of this immunoglobulin also exists in serum. Associated covalently with the polymeric forms of IgA and IgM is a polypeptide, the J chain, which is distinct from both heavy and light chains of immunoglobulins. The importance of this additional polypeptide chain to the integrity of polymeric immunoglobulin structures has been well documented. Furthermore, the presence of the J chain in myeloma cells secreting monomeric IgG (Kaji and Parkhouse, 1974) and in normal IgD and IgG producing cells (Brandtzaeg, 1974) serves to enhance speculation that the J chain has a dominant biological role in humoral immunity and also may be important in the activation of B lymphocytes. Indeed, recent studies have shown a parallel synthesis between J chain and immunoglobulin heavy and light chains during mitogen-induced differentiation of B lymphocytes (Mestecky et al., 1977). In light of these observations we have undertaken a study to elucidate the complete amino acid sequence of J chain, isolated from a human Waldenströms IgM protein. The results described in this and the preceding article provide a complete covalent structure of human J chain.

Materials and Methods

J chain used in these studies was isolated from a Waldenströms IgM macroglobulin (Gray) as previously reported (Mole et al., 1974). Procedures for enzymatic digestion, column chromatography, fractionation of peptides, and amino acid analysis have been described in the preceding paper (Bhown et al., 1977).

Edman Degradations. Automated Edman degradations were performed in a Beckman Model 890C sequencer employing the 1 M Quadrol program developed by Beckman Instruments (022574) and the 0.1 M Quadrol program of Brauer et al. (1975). The latter system was modified slightly for our instrument. The Quadrol concentration was increased to 0.5 M which enhanced the coupling efficiency with PITC.[1] Sequencer reagents and solvents were obtained from Beckman Instruments or Pierce Chemical Co. Dithioerythritol (15 mg/L) was added to the 1-chlorobutane to improve the recoveries of Pth-Ser and Pth-Thr. Repetitive yields for peptides were generally between 92 and 97%.

Manual Edman degradations were conducted using modifications of the three-stage method of Edman (1960). Peptide (10-40 nmol) was first lyophilized in a series of 2.5-mL graduated centrifuge tubes, one for each cycle of the degradation. Coupling was performed in 200 μL of 50% aqueous pyridine (v/v) using 5 μL of PITC. The coupling reaction was carried out under a nitrogen atmosphere, for 30 min, at 50 °C. The aqueous phase, containing the PTC peptide, was then dried, in a vacuum desiccator, over P_2O_5 for 30 min.

Cleavage was performed, under nitrogen, using 200 μL of TFA, for 10 min at 50 °C. If Gln was suspected as the penultimate amino-terminal amino acid, heating was omitted. TFA was then quickly evaporated under vacuum, in the presence of damp NaOH pellets. Following cleavage the residue was redissolved in 200 μL of distilled water and two 1.0-mL butyl acetate extractions were performed and the organic phases discarded.

† From the Division of Clinical Immunology and Rheumatology, Department of Medicine and Department of Microbiology, University of Alabama in Birmingham, Birmingham, Alabama 35294. *Received February 8, 1977.* This work was supported by Grants AI 09153 and AM 03555 from the National Institutes of Health and Grant BMS 74-11522 from the National Science Foundation.

[1] Abbreviations used: PITC, phenyl isothiocyanate; PTC, phenylthiocarbamoyl; TFA, trifluoroacetic acid; Pth, phenylthiohydantoin; butyl-PBD, 2-(4′-*tert*-butylphenyl)-5-(4″-biphenyl)-1,3,4-oxidazole; PCA, 2-pyrrolidone-5-carboxylic acid; PCAase, pyrrolidonecarboxylyl peptidase; DEAE, diethylaminoethyl; TPCK, 1-chloro-4-phenyl-3-tosylamido-2-butanone; CPA, carboxypeptidase A; CPB, carboxypeptidase B; dansyl chloride, 1-dimethylaminonaphthalene-5-sulfonyl chloride; TLC, thin-layer chromatography; GLC, gas-liquid chromatography.

Some peptides were soluble in the butyl acetate layer and an alternate method had to be devised. This second procedure employed three 1.0-mL benzene extractions immediately after coupling with PITC and a drying step was substituted for the butyl acetate extraction.

For both protocols, the residual peptides were analyzed using subtractive or dansyl chloride procedures (Gray, 1967; Woods and Wang, 1967).

Identification of Amino Acid Phenylthiohydantoins. The dried thiazolinone amino acid derivatives maintained overnight in the fraction collector of the sequencer were converted to the phenylthiohydantoin amino acids in 0.2 mL of 1 N HCl, for 10 min at 80 °C (Edman and Begg, 1967). One percent ethanethiol was included in the conversion acid to reduce losses of the labile derivatives of Ser and Thr. Extraction with ethyl acetate (2 × 1.0 mL) was then performed to separate the ethyl acetate soluble Pth-amino acids from those of Arg and His.

The ethyl acetate soluble phenylthiohydantoins were identified and quantitated on a 10% SP-400 support with a Hewlett-Packard Model 5830A gas chromatograph, using an adaptation of the procedure of Hermodson et al. (1972). The instrument was programmed for an initial temperature of 180 °C for 3 min which was increased at a rate of 10 °C per min to 285 °C. This temperature was maintained for an additional 4 min. The temperatures of the inlet port and detector were set for 250 and 300 °C, respectively. In some cases, amino acid analysis following acid hydrolysis of the Pth-amino acid in 6 N HCl, at 130 °C, for 20 h was used. Verification of each amino acid was also obtained by two-dimensional thin-layer chromatography in the presence of butyl-PBD (Summers et al., 1973).

Pth-Arg was identified by the phenanthrenequinone reaction (Yamada and Itano, 1966). Pth-His was identified by the Pauly reaction (Bennett, 1967).

Amino Acid Analysis. Samples for amino acid analysis were hydrolyzed in vacuo in 6 N HCl, at 110 °C, for 20 h. A Durrum D-500 high-pressure amino acid analyzer, controlled by a PDP-8/L computer, was employed for quantitation (Benson, 1972).

Carboxypeptidase Digestion. Carboxypeptidase A and B hydrolyses of individual peptides were performed according to the method of Amber (1972).

Results

Amino-Terminal Sequence (Residues 1–17). The amino terminus of J chain contains a PCA residue which does not react with PITC (Meinke and Spiegelberg, 1972; Mendez et al., 1973a). Thus, a direct approach with the sequencer to secure amino acid sequence data of the intact protein was not feasible. The enzyme, pyrrolidonecarboxylyl peptidase, was found to selectively cleave the PCA residue of carboxyamidomethylated J chain and permitted utilization of the automated sequencer to establish the sequence of this region of the polypeptide chain (Mole et al., 1976).

Amino Acid Sequence of the Tryptic Peptides. Analysis of the fragments generated by extensive trypsin hydrolysis of the J chain (1:50 for 6 h at 37 °C) established the amino acid sequence for 84 of 129 residues in the molecule. In most cases, the sequence was obtained for the entire peptide. Two peptides, Tn and LTP-3-2, encompassing residues 1-5 and 55-94, respectively, were never isolated in these analyses. In order to obtain an entire set of tryptic peptides, citraconyl J chain was subjected to limited hydrolysis with trypsin (1:200 for 1 h at 25 °C) and resulted in the isolation of LTP-3-2 (Bhown et al., 1977). In both procedures, Tn which contains the PCA group was never encountered due to its nonreactivity with fluorescamine. However, this peptide was obtained by paper electrophoresis (Mole et al., 1976). The analysis of all tryptic peptides is presented in the next few paragraphs.

Peptide T-3-4 (Residues 6–17). The amino acid composition of this dodecapeptide contained 2 mol of carboxymethylcysteine per mol. TLC of the dansyl derivative identified N-terminal Ile and a spot migrating as the dipeptide, Ile-Val. Digestion of 10.4 nmol of T-3-4 with carboxypeptidase B for 2 min released 9.2 nmol of Arg. Treatment of 10.4 nmol with a mixture of carboxypeptidases A and B released 10.3 nmol of Arg and 7.6 nmol of Ala in 30 min. The complete sequence was then determined with the sequencer on 290 nmol of T-3-4 as: Ile-Val-Leu-Val-Asp-Asn-Lys-CMCys-Lys-CMCys-Ala-Arg. These results were consistent with those obtained using PCAase-treated J chain and confirmed the location of T-3-4 at the amino terminus of the polypeptide chain. The sequence at positions 15-17 of T-3-4 is identical in composition with a tripeptide isolated by Mendez et al. (1973b) which is disulfide bonded to the C-terminal octapeptide of the α chain in IgA.

Peptide T-6-3 (Residues 18–21). Three steps of the subtractive Edman degradation procedure were sufficient to establish the sequence Ile-Thr-Ser-Arg.

Peptide T-3-1-2 (Residues 22–33). The entire sequence of this very acidic dodecapeptide was established in the sequencer and later confirmed manually with the subtractive Edman method as: Ser-Ser-Glu-Asp-Pro-Asn-Glu-Asp-Glu-Ile-Val-Arg. A second peptide (T-2-4-2) was purified and analyzed through four subtractive Edman degradations as Ser-Ser-Glx-Asx(Pro,Asx,Glx,Asp,Glu,Ile,Val)-Arg and was assumed to represent the same portion of J chain as T-3-1-2.

Peptide T-5-4 (Residues 34–41). The amino acid composition of a 24-h hydrolysate of this peptide indicated 1 mol of Ile per mol. Results of both automated and manual procedures with T-5-4 clearly identified an Ile-Ile linkage which is known to resist the usual conditions of acid hydrolysis. The total sequence established for this octapeptide was Ile-Ile-Val-Pro-Leu-Asp-Asn-Arg.

Peptide T-2-1 (Residues 42–52). The localization of the sole carbohydrate moiety in the J chain was ascertained by analysis of this undecapeptide, the only tryptic peptide which contained significant levels of glucosamine and galactosamine. T-2-1 (430 nmol) was subjected to automated Edman degradation and positive identification of the Pth-amino acids from the sequencer was achieved at each step of the analysis except for the second cycle. Moreover, after two steps of manual Edman degradation, the majority of the carbohydrate and 1 mol each of Glu and Asp were absent in the residual peptide. It is therefore likely that Asn occurs in position 2 of the peptide since glucosamine is linked to the amide nitrogen of Asn in other carbohydrate-containing peptides (Spiro, 1970; Shimigu et al., 1971). On the basis of these data, the sequence of the glycopeptide was deduced as: Glu-Asn(CHO)-Ile-Ser-Asp-Pro-Thr-Ser-Pro-Leu-Arg. Another peptide, T-2-2-1, eluted slightly behind T-2-1 on DEAE-cellulose (Bhown et al., 1977) and was identical in sequence with T-2-1 through the first seven residues. Since J chain is known to contain sialic acid, it is likely that T-2-1 and T-2-2-1 are resolved as a result of carbohydrate heterogeneity in the peptides.

Peptide T-6-4-2 (Residues 53–54). One Edman degradation was sufficient to establish the sequence Thr-Arg for this basic dipeptide.

Peptide LTP-3-2 (Residues 55–94). The largest tryptic peptide encountered, a tetracontapeptide, was obtained only after limited hydrolysis of citraconyl J chain. The peptide

TABLE I: Phenylthiohydantoin Amino Acid Yields Obtained with Peptides Sequenced by Automated Edman Degradation.[a]

Cycle	T-3-4	T-2-1	T-2-4-3	LTP-3-2	CH-4[f]	A-2	A-4[f]	A-1
0	(290)	(430)	(381)	(59)	(78)	(756)	(114)	(52)
1	Ile (224)	Glu (373)	Val (325)	Phe (11)	His[e]	Pro (429)	Pro (91)	Pro (47)
2	Val (319)	Asn[d]	Tyr (108)	Val (46)	Leu (63)	Asn (1100)	Thr[c]	Thr[c]
3	Leu (274)	Ile (400)	Gly (383)	Tyr (61)	Ser (ND)	Glu (618)	Ser[c]	Glu (61)
4	Val (182)	Ser[c]	Gly (329)	His[e]	Asp (41)	Asp (343)	Pro (74)	Val (41)
5	Asp (96)	Asp (280)	Glu (414)	Leu (39)	Leu (67)	Glu (427)	Leu (187)	Glu (52)
6	Asn (106)	Pro (268)	Thr (212)	Ser (24)	CMCys (69)[g]	Ile (408)	Arg (17)	Leu (51)
7	Lys (77)	Thr[c]	Lys (52)	Asp (46)	Lys (42)	Val (449)	Thr (40)[h]	Asp (43)
8	CMCys (131)	Ser[c]	Met (94)	Leu (34)	Lys/Gln (23/37)	Arg[i]	Arg (77)[i]	Asn[c]
9	Lys (6.2)	Pro (170)	Val (25)	CMCys (26)	CMCys (42)[g]	Ile (151)	Phe (112)	Gln[c]
10	CMCys (46)	Leu (168)	Glu (12)	Lys (8.5)	Asp (24)	Ile (185)	Val (113)	Ile (25)
11	Ala (68)	Arg[b]	Thr (ND)	Lys/Gln[c]	Pro (12)	Val[c]	Tyr (74)	Val (29)
12	Arg[b]		Ala (ND)	CMCys (14)	Thr (21)[h]	Pro[c]	His (ND)	Thr[c]
13			Leu (1)	Asp[c]	Glu (14)	Leu[c]	Leu (58)	Ala (30)
14				Pro (11)				Thr[c]
15				Thr[c]				Gln (16)
16				Glu (11)				CMCys (11)
17				Val (19)				Asx[c]
18				Glu (8.2)				Ile (23)
19				Leu (21)				CMCys (28)
20				Asp (11)				Asp (13)
21				Asn[c]				Glu (13)
22				Gln (ND)				Asn (19)
23				Ile (12)				Ser (ND)
24				Val (8.1)				Ala (2.5)
25				Thr (ND)				Ser (ND)
26				Ala (11)				Glu (9.1)
27								Arg (ND)
28								Thr (ND)
29								Tyr (22)
Repetitive yield[j]	88%	90%	97%	91%	92%	88%	99%	95%

[a] Yields of the amino acid phenylthiohydantoins shown in parentheses were determined by GLC except where noted. ND = Not determined. [b] Qualitative identification was made after hydrolysis of aqueous layer. [c] Qualitative identification was made by TLC. [d] Confirmed by subtractive Edman degradation; carbohydrate attached. [e] Qualitative identification was made by the Pauly reaction. [f] Quantitated on an amino acid analyzer after acid hydrolysis of ethyl acetate or aqueous layers. [g] Quantitated as alanine after acid hydrolysis. [h] Quantitated as α-aminobutyric acid after acid hydrolysis. [i] Identification confirmed by the phenanthrenequinone reaction. [j] Calculated as follows: T-3-4, Val-2 to Val-4; T-2-1, Ile-3 to Leu-10; T-2-4-3, Val-1 to Val-9; LTP-3-2, Val-2 to Val-17; CH-4, A-2, A-4, and A-1 from their regression curves.

contained an exceptional number of Asp and Glu residues as well as the sole Phe and His residues in the J chain. LTP-3-2 (59 nmol) was subjected to 26 successive cycles with the sequencer and identification of 24 phenylthiohydantoin derivatives was achieved by GLC and TLC. These results are presented in Table I. To illustrate repetitive yield, these results are also plotted in Figure 1. The sequence determined was Phe-Val-Tyr-His-Leu-Ser-Asp-Leu-CMCys-Lys($\frac{\text{Lys}}{\text{Gln}}$)-CMCys-Asp-Pro-Thr-Glu-Val-Glu-Leu-Asp-Asn(Gln)-Ile-Val(Thr)-Ala(Thr,Gln,CMCys,Asx,Ile,CMCys,Asp,Glu,-Asn,Ser,Ala,Ser,Glu,Arg).

Peptide T-6-2 (Residues 95–98). Both the dansyl Edman and the subtractive Edman procedures showed this tetrapeptide had the sequence Thr-Tyr-Asx-Arg. The electrophoretic mobility of T-6-2 at pH 6.5 indicated that Asn, rather than Asp, was present in the peptide and Asn was confirmed later with the sequencer.

Peptide T-4-5 (Residues 99–108). Although compositional analysis indicated T-4-5 resulted from incomplete modification of Lys with citraconic anhydride, residual chymotrypsin activity present in TPCK-treated trypsin preparations was probably responsible for its isolation in high yield. Nine cycles of the subtractive degradation revealed the sequence Asx-Lys-CMCys-Tyr-Thr-Ala-Val-Val-Pro-Leu. Asparagine was detected by GLC as the amino terminus of T-4-5; however, the sample was extracted from the sequencer cup during the fourth cycle, due to the hydrophobicity of the residual peptide.

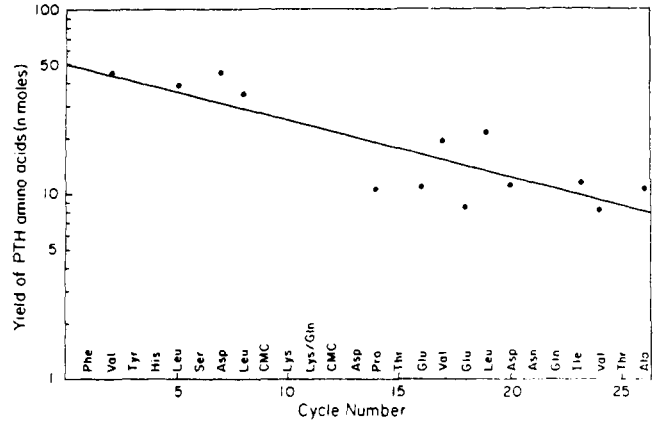

FIGURE 1: Semilog plot of yields of Pth-amino acids from a sequencer run of LTP-3-2. The sequence shown below the graph was ascertained from GLC, TLC, and by amino acid analysis after back hydrolysis.

Peptide T-2-4-3 (Residues 109–129). Thirteen automated Edman degradations with 381 nmol established the partial sequence of T-2-4-3 as Val-Tyr-Gly-Gly-Glu-Thr-Lys-Met-Val-Glu(Thr,Ala)-Leu. Yields fell sharply at Lys-115. Two additional tryptic peptides T-2-2-2 and T-2-4-1 derived from this same region of the J chain were also isolated and sequenced. By subtractive analysis, T-2-2-2 gave Val-Tyr-Gly-

Gly(Glu,Thr)-Lys and T-2-4-1 gave Met-Val-Glx-Thr-Ala-Leu-Thr-Pro-Asx-Ala-CMCys-Tyr-Asx. These sets of data corroborate our earlier studies with a cyanogen bromide peptide (Mole et al., 1974) and furnish an overlap at Met-116 of the polypeptide.

Amino Acid Sequence of Selected Chymotryptic Peptides. To compile a continuous sequence of the tryptic peptides, we attempted to isolate and analyze peptides from chymotrypsin digestion. This approach proved cumbersome since extensive fragmentation of the J chain occurred with chymotrypsin and produced a multitude of fragments, most of which were extremely difficult to purify. For this reason, we focused our attention on the arginine-containing fractions from chymotrypsin hydrolysis since all of the tryptic peptides except T-4-5 possessed a C-terminal Arg residue. The following paragraphs present the sequence data obtained from these chymotryptic peptides.

Peptide C-3-6 (Residues 9–15). This heptapeptide was obtained in sufficiently high yields for analysis and was found to contain 2 mol each of carboxymethylcysteine and Lys. The sequence Val-Asx-Asx-Lys was adequate to establish its position within the amino terminus of the J chain.

Peptide C-5-5 (Residues 16–19). The sequence of this basic tetrapeptide was ascertained by the subtractive Edman procedure as Ala-Arg-Ile-Thr and furnished an overlap between tryptic peptides T-3-4 and T-6-3.

Peptide C-2-1 (Residues 39–51). C-2-1 was similar in composition to the subtilisin glycopeptide isolated by Meinke and Spiegelberg (1972) but differs by an additional Pro and Asp. Since both are glycopeptides, they most probably were derived from the same segment of the J chain. Subtractive Edman degradation of C-2-1 revealed the sequence Asx-Asx-Arg-Glx-Asx-Ile and established an overlap between tryptic peptides T-5-4 and T-2-1.

Peptide CH-7 (Residues 52–55). This tetrapeptide was found to contain impurities. Nevertheless, the sequence Arg-Thr-Arg-Phe was determined by the subtractive Edman procedure. Furthermore, treatment of CH-7 with CPA released only Phe. Hydrolysis with a mixture of CPA and CPB released Phe and Arg. Inasmuch as J chain contains one Phe residue per mol, it was evident that T-2-1, T-6-4-2, and LTP-3-2 were linked, in that order.

Peptide CH-4 (Residues 58–96). Automated sequencer analysis of this large peptide for 13 cycles was analogous to the results obtained with tryptic peptide LTP-3-2. The sequence was determined with 78 nmol as His-Leu(Ser)-Asp-Leu-CMCys-Lys-($^{Lys}_{Glu}$)-CMCys-Asp-Pro-Thr-Glu and confirmed the presence of both Lys and Gln in position 65 of the J chain sequence. The quantity of lysine obtained was significantly greater than the carry over of previous cycles of the degradation.

Peptide C-4-9 (Residues 97–102). Manual Edman analysis of C-4-9 gave the complete sequence Asx-Arg-Asx-Lys-CMCys(Tyr). Since Asp precedes Arg only once in the J chain sequence, the data permitted the alignment of T-6-2 with T-4-5.

Peptide CH-1-3 (Residues 103–110). The timely acquisition of this chymotryptic peptide furnished an overlap between T-4-5 and T-2-4-3. The peptide bond between Leu-108 and Val-109 of the J chain was easily cleaved during trypsin digestions, even though the precaution of using TPCK-treated trypsin was considered. CH-1-3 was observed only under limiting conditions of hydrolysis with chymotrypsin (Bhown et al., 1977). Manual degradation provided the entire sequence of CH-1-3 as Thr-Ala-Val-Val-Pro-Leu-Val-Tyr.

Peptide C-3-4 (Residues 111–119). The first four amino acids of this nonapeptide were determined by subtractive Edman degradation as Gly-Gly-Glx-Thr(Lys,Met,-Val,Glx,Thr). A second Met containing peptide, C-4-6, was obtained in low yield but gave the sequence Gly-Gly-Glx-Thr-Lys(Met). Therefore, C-4-6 undoubtedly resulted from further cleavage of C-3-4.

Amino Acid Sequence of Peptides Derived from Cleavage of Aspartylprolyl Peptide Bonds. A preliminary study of the isolation, alignment, and partial structural analysis of peptides generated by acid-catalyzed hydrolysis of aspartylprolyl linkages of the J chain has been published (Mole et al., 1977). Since that report was made, an extension of the sequence data on these fragments has been obtained and is briefly summarized below.

Peptide A-3 (Residues 1–25). A-3 was the sole fragment isolated from acidic hydrolysis of J chain that failed to react with PITC or dansyl chloride. Asp could be detected after 8 h of incubation with CPA at pH 5.0 and 37 °C but was not released when digestions were conducted at pH 7.0. To further evaluate the origin of A-3, the peptide was reacted with citraconic anhydride and digested with trypsin, and the peptides were separated by electrophoresis at pH 3.65. One ninhydrin-negative peptide (A-3-2-1) and a ninhydrin-positive peptide (A-3-2-2) were recovered and on analysis gave the following amino acid composition: $Asp_{1.1}$, $Arg_{0.9}$, $Glu_{3.0}$ (A-3-2-1); $Asp_{1.0}$, $Ser_{1.8}$, $Glu_{0.5}$ (A-3-2-2). Although the quantity of sample prohibited the sequence determination of these peptides, the data demonstrate that peptide A-3-2-1 was identical with the five N-terminal amino acids of the J chain (Mole et al., 1976) and that peptide A-3-2-2 represented the amino-terminal four residues of tryptic peptide T-3-1-2.

Peptide A-2 (Residues 26–46). Automated Edman degradation on 756 nmol of this glycopeptide was performed through 13 repetitive cycles and gave the sequence Pro-Asn-Glu-Asp-Glu-Ile-Val-Arg-Ile-Ile-Val-Pro-Leu. Tryptic peptide T-3-1-2 is therefore juxtaposed with peptide T-5-4 in the intact polypeptide chain.

Peptide A-4 (Residues 47–67). No additional data from that already reported were obtained on this peptide (Mole et al., 1977). The sequence ascertained was Pro-Thr-Ser-Pro-Leu-Arg-Thr-Arg-Phe-Val-Tyr(His)-Leu(Lys_2,$CMCys_2$,-Asx_2,Ser,Glx,Leu).

Peptide A-1 (Residues 68–129). Four separate sequencer runs on this peptide were necessary to determine the sequence Pro-Thr-Glu-Val-Glu-Leu-Asp-Asn-Gln-Ile-Val-Thr-Ala-Thr-Gln-CMCys-Asx-Ile-CMCys-Asp-Glu-Asn(Ser)-Ala-(Ser)-Glu(Arg,Thr)-Tyr(Asp,Arg,Asn,Lys,CMCys)-Tyr-(Thr,Ala,Val)-Val. Yields for individual Pth-amino acids of one of these runs are shown in Figure 2 and Table I. Difficulties in the identification of the Pth derivatives of Ser, Thr, Asp, Lys, and carboxymethylcysteine occurred well into the sequence of the peptide due to the low GLC responses of those Pth-amino acids. Nevertheless, Tyr at positions 29 and 35 and a Val at position 39 were sufficient to permit alignment of tryptic peptides LTP-3-2 with T-6-2. The Ser and Arg residues assigned to positions 23, 25, and 27 of A-1 were deduced from the amino acid compositions of peptides A-1, LTP-3-2, and CH-4.

Discussion

The most pragmatic approaches to the determination of protein primary structure capitalize on direct sequencer analyses of the intact molecule and of the relatively few fragments generated through the cyanogen bromide cleavage reaction. These procedures are advantageous in that large polypeptides readily adapt to automated methods of analysis and

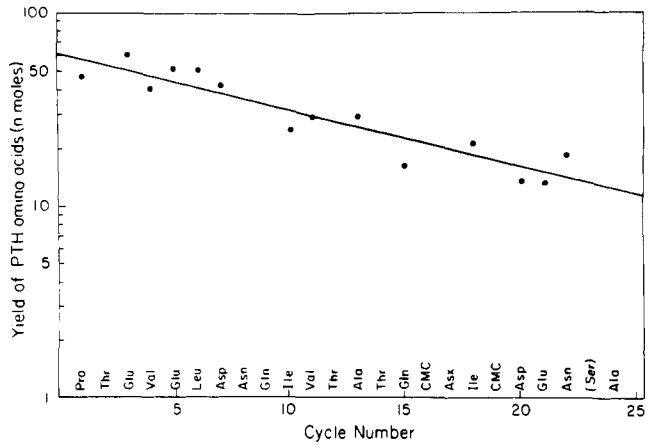

FIGURE 2: Semilog plot of yields of Pth-amino acids from a sequencer run of A-1. The sequence deduced from this experiment and three other runs is presented below the graph.

yield considerable structural data with little effort. The existence of an N-terminal PCA residue and the location of a single methionine residue 14 amino acids from the C terminus in the J chain moderate to a large extent the technical approaches which can be applied to the structural definition of this molecule. Thus several chemical and enzymatic cleavage procedures were necessary to attain a complete amino acid sequence. These included: (1) cyanogen bromide cleavage and separation of the C-terminal and N-terminal fragments; (2) controlled enzymatic hydrolysis with trypsin and chymotrypsin employing limited and prolonged incubation conditions followed by fractionation of the peptides from these digestions; (3) cleavage of the cyclized N-terminal glutamyl residue with PCAase which enabled direct sequence analysis on the intact J chain molecule; and (4) selective acid hydrolysis of aspartylprolyl peptide bonds and isolation of the four subsequent fragments.

Tryptic peptides provided the sequence for 108 of 129 amino acids or roughly 85% of the entire structure. To secure overlapping sequences several attempts at producing chymotryptic peptides were made. For the most part, however, chymotrypsin digestions were disappointing. Chymotrypsin-catalyzed hydrolyses occurred at multiple sites and the result was extensive degradation of the J chain. Nevertheless, we were able to purify five chymotryptic peptides which permitted the ordering of nine tryptic fragments.

The high content of Pro and Asp obtained in compositional analyses of J chain suggested aspartylprolyl cleavage might be profitable particularly in view of the sequence information already on hand. Our initial work with the tryptic peptides confirmed the occurrence of three aspartylprolyl peptide bonds and encouraged us to investigate low-temperature acid-catalyzed hydrolysis.

The choice of acid hydrolysis as a means of fragmentation was fortunate. Acidic cleavage at aspartylprolyl peptide bonds was essentially quantitative and the four peptides, designated A-1, A-2, A-3 and A-4, were easily purified in high yield by gel filtration (Mole et al., 1977). A substantial portion of the sequence of three of these peptides was readily achieved with the automatic sequencer. Only with peptide A-3, the N-terminal peptide containing the unreactive PCA residue, was sequence data unobtainable.

From the data summarized in Figure 3, it is possible to construct the amino acid sequence of human J chain. The sum of the integral numbers of amino acids calculated from the sequence shown in Figure 3 is virtually identical with the amino acid composition of whole J chain (Mestecky et al., 1971). Furthermore, the molecular weight of 16 422 calculated from our data, and corrected to include 7.6% carbohydrate, agrees well with the experimentally determined value of 15 600 obtained by sedimentation equilibrium studies (O'Daly and Cebra, 1971; Schrohenloher et al., 1973; Wilde and Koshland, 1973).

The exact number of Cys residues in the J chain has been uncertain, due to the difficulties in the quantitation of carboxymethylcysteine. However, both our analytical data and the results of sequence determination establish the number as 8. Data available pertaining to the sites of J chain attachment to immunoglobulin suggest that Cys-13 and -15 are joined through interchain disulfide bridges with the immunoglobulin heavy chain. First, a peptide identical in amino acid composition with the sequence at position 15–17 of the J chain was demonstrated to be disulfide bonded to the penultimate Cys residue in IgA (Mendez et al., 1973b). Secondly, a large N-terminal fragment of the J chain, containing two Cys residues, is released from subtilisin-treated IgM only after reduction of the disulfide bonds (Koshland et al., 1977). The implication is that both Cys residues are linked to heavy chains, since J chain does not contain a free Cys residue (Chapius and Koshland, 1974). Moreover, an intrachain disulfide bond exists between Cys residues in position 126 and an unidentified site in the J chain sequence (Mestecky and Schrohenloher, 1974; Mestecky et al., 1974). The remaining four Cys are probably linked via intrachain disulfide bridges, although solid data to support this conclusion are presently unavailable. Assignment of the disulfide arrangements has been difficult due to the necessity of cleaving all disulfide bonds in the preparation of J chain. Lack of aspartylprolyl bonds in the Fc portion of IgM (Putnam et al., 1973) suggests that low-temperature acid hydrolysis might be a useful alternative to obtain J chain with the disulfide bonds intact. It will be important in understanding the conformation of J chain to ascertain whether bridges are located between adjacent Cys residues or between distant Cys residues.

Human J chain contains 7.6% total carbohydrate by weight (Niedermeier et al., 1972). Investigations herein have demonstrated that all of the oligosaccharide is linked to an Asn at position 43. Evidence to support a single location of carbohydrate in the primary structure is fairly conclusive. First, the glycopeptides isolated from trypsin, chymotrypsin, and acid hydrolysis were derived from the same region of the J chain. Secondly, subtractive Edman degradation of tryptic peptide T-2-1 resulted in simultaneous loss of Asp and the amino sugars after the second cycle. Thirdly, the sequence Asn-Ile-Ser is characteristic of the sequence Asn-X-$^{Thr}_{Ser}$, in other proteins, where the carbohydrate moiety is linked via an Asn residue (Spiro, 1970). Other investigators (Meinke and Spiegelberg, 1972) have isolated glycopeptides from subtilisin digestions of J chain of IgA and IgM. These peptides were similar in composition to a chymotryptic peptide (C-2-1) isolated in the present investigation. On the basis of their studies, they concluded that the oligosaccharide was located at an Asp or Asn corresponding to residues 39 or 40 of our sequence. Since the tryptic glycopeptide T-2-1 in our studies was devoid of these amino acids, it is unlikely that the oligosaccharide linkage site occurs at either of these positions. It is noteworthy that the oligosaccharide structure is confined to a portion of the J chain apparently free of disulfide bridges (residues 16–62). This is undoubtedly significant since the carbohydrate is hydrophilic and is likely to be located at the surface of the molecule. Furthermore, it is reasonable to assume the bulky carbohydrate structure might greatly affect the conformation of the poly-

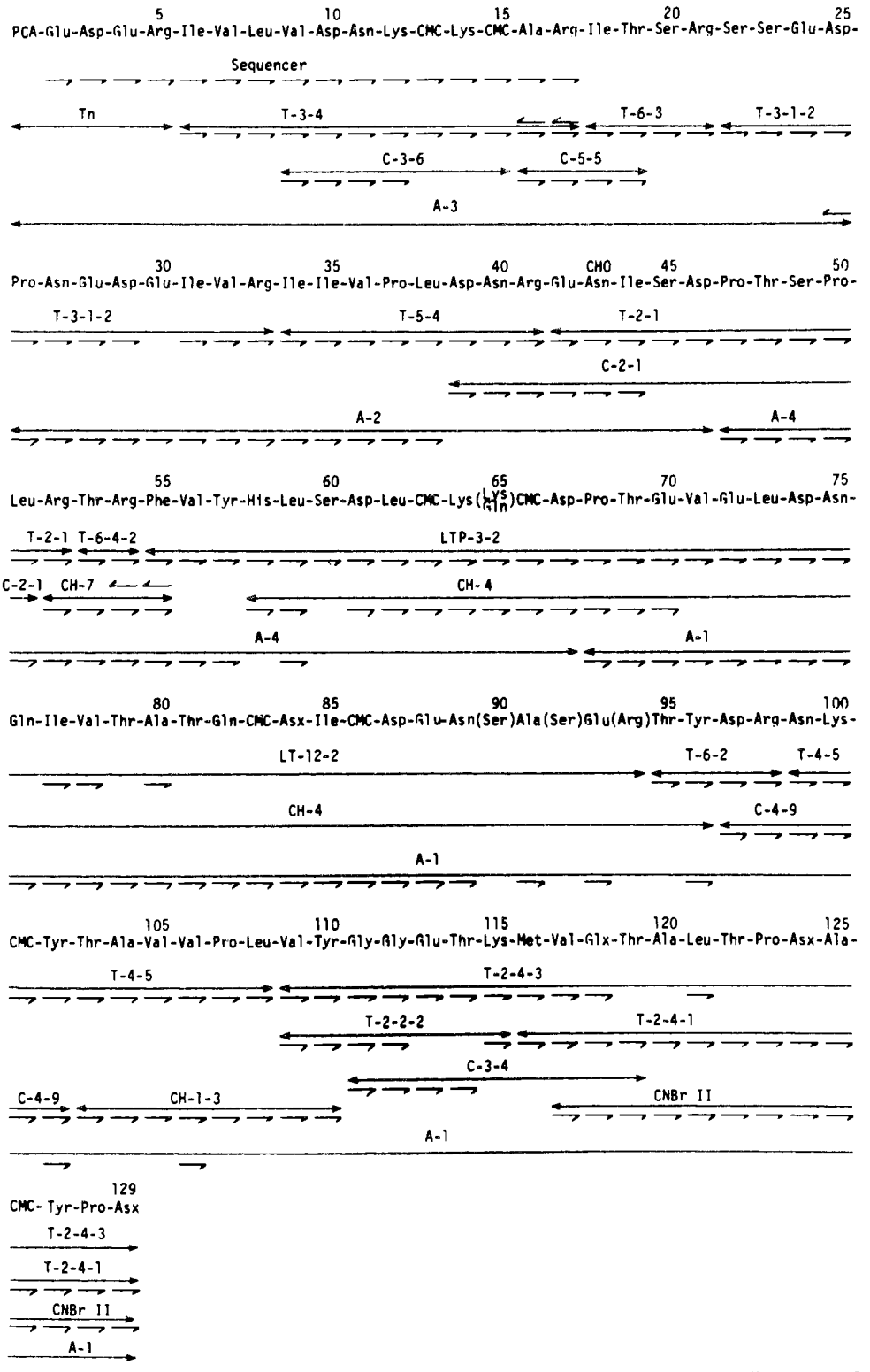

FIGURE 3: The amino acid sequence of human J chain. Peptides from tryptic, chymotryptic, and acid hydrolyses are indicated by T, C, and A, respectively. Peptides from limited trypsin and chymotrypsin hydrolyses are designated LTP and CH, respectively. CNBr indicates a peptide obtained from cyanogen bromide cleavage. Other symbols used are: → for residues identified by the Edman degradation: ← for residues identified by digestion with carboxypeptidases (A or B).

peptide chain and thus the interaction with immunoglobulin.

An interesting feature of the primary structure of J chain is the presence of distinctly hydrophobic regions (positions 6–9, 34–38, and 104–109). Although the significance of these areas is not immediately obvious, it is not unreasonable to assume these to be important to the conformation of the polypeptide chain. Four of the eight Pro residues are found in a stretch of 25 amino acids (residues 26–50) that lack disulfide bonds but does contain the carbohydrate moiety. Although this region probably does not correspond to the hinge region of immunoglobulins, the restrictions normally imposed by disulfide bridges do suggest some flexibility of the tertiary structure in this region of the polypeptide chain. Interestingly, the aromatic amino acids are positioned in the latter one half of the sequence, and most occur in the C-terminal one-quarter of the

structure. Since aromatic amino acids are important in protein-protein interaction mechanisms, it is tempting to speculate that this region plays some important role in the function of J chain, perhaps in its structural relationship to immunoglobulin.

The identification of Lys and Gln at position 65 of the sequence was unexpected, inasmuch as the J chain used was obtained from a single Waldenström patient. Both amino acids were found in tryptic (LTP-3-2) and chymotryptic (CH-4) peptides and therefore do not reflect technical difficulties in our analyses. Two hypotheses for the apparent heterogeneity in the J chain can be advanced. First, the genes responsible for J chain synthesis do not demonstrate allelic exclusion. This explanation directly contrasts the situation known to exist in the synthesis of immunoglobulin polypeptide chains and is probably unlikely. Alternatively, a point mutation of the plasma cell line prior to proliferation could account for the heterogeneity observed. The latter conclusion is supported by the observation that a single mutation in the codon for lysine (AAA or AAG) would then result in the incorporation of glutamine (CAA or CAG). Certainly before any firm conclusions can be made regarding the immunogenetics of the system, similar heterogeneity in the J polypeptides associated with other myeloma proteins must be clearly established.

Our sequence data were further evaluated through the National Biomedical Research Foundation for possible matches with sequences of other polypeptide chains. A computer search of five segments comprising the entire J chain sequence was compared with 59 674 test segments of equal length. Unexpectedly, the search failed to indicate significant sequence homology with over 200 different proteins including the immunoglobulin heavy and light chains.

J chains have now been isolated from several different animal sources, among them mouse (Rosenstein and Jackson, 1973; Barger and Inman, 1976), rabbit (O'Daley and Cebra, 1971), dog (Kehoe et al., 1972), pig (Zikan, 1973), and nurse shark (McCumber and Clem, 1976). Completion of the amino acid sequence of human J chain reported here will facilitate comparisons of the sequence with that determined for other species and should provide a basis for assessing common structural features relative to the function of J chains and their interaction with polymeric immunoglobulins.

Acknowledgment

The authors are indebted to Mr. Eddie Finch and Mr. Arthur Weissinger for performing the automated sequencer analyses.

Supplementary Material Available

Additional information regarding amino acid compositions and sequence data for chymotryptic and tryptic peptides analyzed by subtractive Edman procedures (5 pages). Ordering information is given on any current masthead page.

References

Amber, R. P. (1972), *Methods Enzymol. 25B,* 143.
Barger, B. O., and Inman, F. P. (1976), *Immunochemistry 13,* 165.
Bennett, J. C. (1967), *Methods Enzymol. 11,* 330.
Benson, J. R. (1972), *Am. Lab. 4,* 53.
Bhown, A. S., Mole, J. E., and Bennett, J. C. (1977), *Biochemistry* (preceding paper in this issue).
Brandtzaeg, P. (1974), *Nature (London) 252,* 418.
Brauer, A. W., Margolies, M. N., and Haber, E. (1975), *Biochemistry 14,* 3029.
Chapius, R. M., and Koshland, M. E. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 657.
Edman, P. (1960), *Ann. N.Y. Acad. Sci. 88,* 602.
Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1,* 80.
Gray, W. R. (1967), *Methods Enzymol. 11,* 139.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11,* 4493.
Kaji, H., and Parkhouse, R. M. E. (1974), *Nature (London) 249,* 45.
Kehoe, J. M., Tomasi, T. B., Ellouz, F., and Capra, J. D. (1972), *J. Immunol. 109,* 59.
Koshland, M. E., Chapius, R. M., Recht, B., and Brown, J. C. (1977), *J. Immunol. 118,* 775.
McCumber, L. J., and Clem, L. W. (1976), *Immunochemistry 13,* 479.
Meinke, G. C., and Spiegelberg, H. L. (1972), *J. Immunol. 109,* 903.
Mendez, E., Frangione, B., and Franklin, E. C. (1973a), *Biochemistry 12,* 1124.
Mendez, E., Prelli, F., Frangione, B., and Franklin, E. C. (1973b), *Biochem. Biophys. Res. Commun. 55,* 1291.
Mestecky, J., and Schrohenloher, R. E. (1974), *Nature (London) 249,* 650.
Mestecky, J., Schrohenloher, R. E., Kulhavy, R., Wright, G. P., and Tomana, M. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 544.
Mestecky, J., Winchester, R. J., Hoffman, T., and Kunkel, H. G. (1977), *J. Exp. Med. 145,* 760.
Mestecky, J., Zikan, J., and Butler, W. T. (1971), *Science 171,* 1163.
Mole, J. E., Bhown, A. S., and Bennett, J. C. (1974), *Biochem. Biophys. Res. Commun. 61,* 457.
Mole, J. E., Bhown, A. S., and Bennett, J. C. (1976), *Biochem. Biophys. Res. Commun. 73,* 92.
Mole, J. E., Bhown, A. S., and Bennett, J. C. (1977), *J. Immunol. 118,* 67.
Niedermeier, W., Tomana, M., and Mestecky, J. (1972), *Biochim. Biophys. Acta 257,* 527.
O'Daly, J. A., and Cebra, J. J. (1971), *Biochemistry 10,* 3843.
Putnam, F. W., Florent, G., Paul, C., Shinoda, T., and Shimigu, A. (1973), *Science 182,* 287.
Rosenstein, R. W., and Jackson, P. (1973), *Biochemistry 8,* 1659.
Schrohenloher, R. E., Mestecky, J., and Stanton, T. H. (1973), *Biochim. Biophys. Acta 295,* 576.
Shimigu, A., Putnam, F. W., Paul, C., Clamp, J. R., and Johnson, I. (1971), *Nature (London), New Biol. 231,* 73.
Spiro, R. G. (1970), *Annu. Rev. Biochem. 39,* 599.
Summers, M. R., Smythers, G. W., and Oroszlan, S. (1973), *Anal. Biochem. 53,* 624.
Wilde, C. E., III, and Koshland, M. E. (1973), *Biochemistry 12,* 3218.
Woods, K. R., and Wang, K. T. (1967), *Biochim. Biophys. Acta 133,* 369.
Yamada, S., and Itano, H. A. (1966), *Biochim. Biophys. Acta 130,* 538.
Zikan, J. (1973), *Immunochemistry 10,* 351.