PARTIAL SEQUENCE OF THE PRECURSORS OF IMMUNOGLOBULIN LIGHT-CHAINS OF DIFFERENT SUBGROUPS: EVIDENCE THAT THE IMMUNOGLOBULIN VARIABLE-REGION GENE IS LARGER THAN HITHERTO KNOWN

Israel Schechter* and Yigal Burstein+
Departments of Chemical Immunology* and Organic Chemistry+, The Weizmann Institute of Science, Rehovot, Israel

Received November 13,1975

SUMMARY. The proteins programmed in the wheat germ cell-free system by the M-41 L-chain mRNA were labeled with [^{35}S]Met, [^{3}H]Leu, [^{3}H]Ile or [^{3}H]Pro, and were subjected to amino acid sequence analyses. The results showed that this mRNA directs the synthesis of two precursors in which 22 or 20 residues precede the N-terminus of the mature M-41 L-chain. Partial sequence of the extra pieces in these precursors are:

Met * Met * * Pro * * Ile * * * Leu Leu Leu Leu * Pro * * * *} Mature M-41
Met * * Pro * * Ile * * * Leu Leu Leu Leu * Pro * * * *} light-chain

Considering the role of Met in the initiation of protein synthesis in eukaryotes, that both precursors contain N-terminal Met, the proximity of the two methionines, and the complete sequence homology, it is reasonable to assume that the two precursors originated from the initiation of M-41 mRNA translation at two points (Met-1 and Met-3 codons).
The extra piece in the precursors is coupled to the variable(V)-region of the L-chains. The precursors of M-41 and M-321 L-chains (which are of different subgroups) contain extra pieces which differ in size and sequence. These findings imply that the extra piece is part of the V-region, i.e., that the gene coding for the immunoglobulin V-region may be larger than hitherto known.

The precursors of immunoglobulin (Ig) light-chains contain extra pieces which are coupled to the variable(V)-regions of the mature proteins (1-4). These findings immediately raised the question whether the extra piece is part of the V-region or whether it represents a new constant polypeptide segment of the Ig chain. To resolve this issue it is necessary to determine the sequence of L-chain precursors of difference subgroups. The M-321 and M-63 L-chains are of the same subgroup, the M-41 L-chain is of a different subgroup (5). The size and partial sequence of the N-terminal extra piece in the precursors of M-321 (1-3) and M-63 (4) L-chains have been reported. In the present study we determine the size and partial sequence of the extra piece in the M-41 L-chain precursor.

Materials and Methods. [^{3}H]Leu (54 Ci/mmole), [^{3}H]Pro (43 Ci/mmole), [^{3}H]Ile (26 Ci/mmole) and [^{35}S]Met (170 Ci/mmol) were purchased from the Radiochemical Centre, Amersham. The ten [^{14}C]-labeled amino acids (Ala,Arg,Gly,Ile,Leu,Lys,Pro,Ser,Thr, Val, 90-320 Ci/mol) were obtained from New England Nuclear. The MOPC-41 (M-41) myeloma (gift of Dr. M. Potter) was maintained as a solid tumor in female BALB/c mice. The mRNA coding for the M-41 L-chain was prepared from M-41 myeloma polysomes specifically precipitated by antibodies to M-41 L-chain, followed by chromatography on oligo(dT)-cellulose (6). The M-321 L-chain mRNA, prepared by the same immune-precipitation procedure, was previously described (1). Translation of the mRNAs in the wheat germ cell-free system, and amino acid sequence analyses of the radioactively

labeled cell-free products in the Beckman Model 890C Automatic Sequencer, have been detailed elsewhere (3). All samples were sequenced twice, and the pattern of radioactive peaks in the duplicate was identical. Repetitive yields were 91-94%. Absolute yields were calculated after correction for the background and "out of step" radioactivity (7). The absolute yields for M-41 labeled precursors ranged between 63 and 77%, with M-321 labeled precursors the yields were 95-106% (see ref. 3).

RESULTS AND DISCUSSION

Identification of Cell-Free Products Programmed by the M-41 mRNA. The fingerprints of tryptic digests of the total cell-free products and of the authentic M-41 L-chain are quite similar (Fig. 1), thus showing that the M-41 mRNA is translated with fidelity. By matching the autoradiagram with the ninhydrin-stained finger-

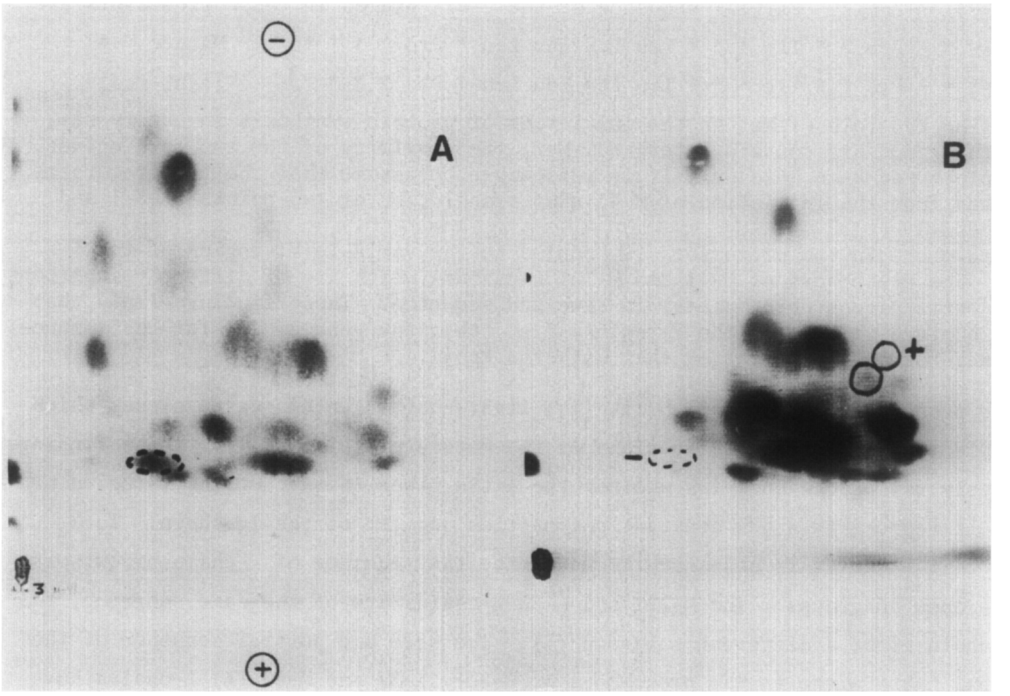

Fig. 1. Tryptic fingerprints of authentic M-41 L-chain (A) and of cell-free products (B) programmed by the M-41 L-chain mRNA in the wheat germ cell-free system. The total reaction mixture labeled with ten [^{14}C]-labeled amino acids was supplemented with 3 mg of M-41 L-chain, was reduced, aminoethylated (9), and digested with tosyl-phenylalanine-chlormethyl ketone treated trypsin (E:S ratio of 1:65) for 12 hr at room temperature. The freeze-dried peptide mixture was loaded on a Whatman 3MM paper and subjected to descending chromatography (1-butanol:acetic acid:pyridine: water, 15 : 3 : 10 : 12 :, 14 hr) followed by high voltage electrophoresis at pH 3.5 (3000 V/55 min). After exposure to X-ray film the paper was stained with ninhydrin. Spots in the autoradiogram and stained paper were matched. The peptide encircled by a dashed line in the fingerprint of the mature protein (A) is missing in the cell-free product (B). Peptides encircled by a continuous line in the cell-free product (B) are missing in the mature protein (A). All other peptides were matched when the original ninhydrin-stained paper and X-ray film were inspected.

print of the mature L-chain, it is seen that only one L-chain peptide is missing, and that two, perhaps three, additional peptides are present in the cell-free digest. The missing peptide probably corresponds to the N-terminal peptide of the mature protein which is modified in the precursor. The additional peptides might come from the extra piece, as well as from new tryptic peptides generated at the C-terminus as a result of incomplete synthesis of the L-chain (see below). Gel electrophoresis of the total cell-free products shows several protein bands (Fig. 2), neither of which is of the same size as the mature M-41 L-chain (23,760 daltons, from ref. 8). The molecular weight of the major component is about 25,000, and the smaller proteins range in size down to about 15,000. The data quoted below suggest that these proteins represent precursor molecules which share identical N-terminus, consequently, the smaller proteins presumably lack portions of the C-terminus.

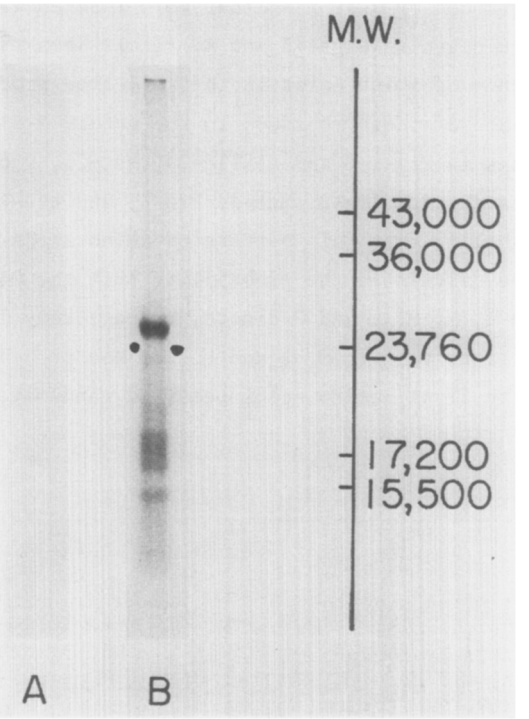

Fig. 2. Autoradiogram of sodium dodecyl sulfate polyacrylamide gel of the proteins programmed by the M-41 mRNA in the wheat germ cell-free system. Total reaction mixtures labeled with ten [^{14}C]-labeled amino acids were reduced and analysed on slab gels (12% acrylamide, 11 V/cm, 3 hr) as described (10). Results from reactions without (A) and with (B) added mRNA are given. Dots indicate the position of mature M-41 L-chain applied with the cell-free sample. Molecular weight standards were ovalbumin, glyceraldehyde phosphate dehydrogenase, M-41 L-chain, myoglobin, and hemoglobin.

The total cell-free products labeled with [^{3}H]Pro and [^{3}H]Ile were subjected to preparative gel electrophoresis (2), and peaks of radioactive material with electrophoretic mobility of about 25,000 and 15,500 daltons were eluted. The patterns of radioactivity obtained from sequencer runs of the total cell-free products and of the two resolved peaks were identical, thus demonstrating that the three samples share identical N-terminus. Results from analyses of the [^{3}H]Pro labeled samples are given in Fig. 3, similar results were obtained with the [^{3}H]Ile labeled materials. These experiments suggest that the size heterogeneity of the cell-free products is due to the initiation of mRNA translation at one major point, with premature termination at several points. A similar situation has been described in the translation of the M-321 L-chain mRNA in the Krebs II ascites cell-free system, where 5 precursor molecules ranging in size from 28,700 to 17,200 daltons were isolated, and all of them were found to share identical N-terminus (2).

Sequence Analyses of M-41 L-Chain Precursors. Results of sequencer runs of the total cell-free products programmed by M-41 mRNA that were labeled with [^{3}H]Pro, [^{3}H]Ile, [^{3}H]Leu, or [^{35}S]Met are given in Figs. 3 and 4. It is seen that after 22 degradation steps, the sequence of amino acids in the cell-free product show perfect homology with the sequence of these residues in the mature M-41 L-chain. That is, (referring to major peaks in Figs. 3 and 4) Ile-24, Met-26, Pro-30 and Leu-33 in the cell-free product match with Ile-2, Met-4, Pro-8, and Leu-11 in the mature M-41 L-chain (8). The probability that this pattern of amino acids occurs by chance is practically nil (we have calculated the probability that the OIOMOOOPOOLOOO sequence occurs by chance, where I, M, P, L and O denote, respectively Ile, Met, Pro, Leu, and amino acid residue other than I, M, P and L. If the probability of I, M, P and L is 0.05, as it would be if all amino acids occurred randomly, the result is:

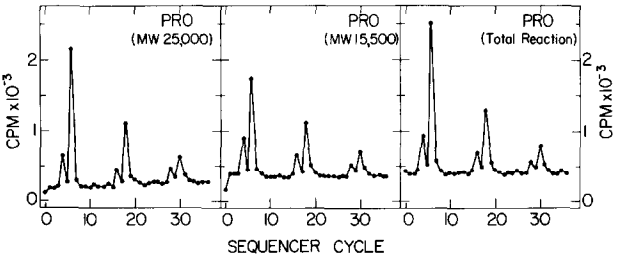

Fig. 3. Radioactivity recovered at each sequencer cycle from cell-free products programmed by the M-41 mRNA, and labeled with [^{3}H]Pro. The results are from analyses of the total cell-free-products, and of the fractions with molecular weight of 25,000 and 15,500 resolved from the total cell-free product by gel electrophoresis. Cycle zero represents a blank cycle (without phenylisothiocyanate) which was used to wash out potential radioactive contaminants.

6.2×10^{-9}). These results establish that the M-41 L-chain mRNA programs the synthesis of a precursor in which 22 amino acid residues are coupled to the N-terminus of the mature protein. From the data given in Figs. 3 and 4 it is seen that the extra piece contains two methionine residues at positions 1 and 3; two prolines at positions 6 and 18; one isoleucine at position 9; and a quadriplet of leucines at positions 13, 14, 15 and 16 (Fig. 5, up). The data also show that minor radioactive peaks precede each of the major peaks by two sequencer cycles. The minor peaks contain about 25% (range 20-30%) of the counts in both the major and minor peaks. Evidently, the pattern of Leu radioactivity at cycles 11-16 is the composite of minor peaks at cycles 11 - 14 and major peaks at cycles 13-16. The Met radioactivity at cycle 1 is the sum of counts from Met-1 and from the minor peak which precedes Met-3. Perfect sequence homology between the major and minor peaks is obtained by displacement of the minor peaks by two residues (Fig. 7). It is thus apparent that about 25% of the precursor molecules are shorter by two residues at the N-terminus (Figs. 5,7).

The two M-41 precursors could originate from 2 distinct mRNA molecules which differ by 2 codons. Alternatively, there may be one mRNA species containing the information for the larger precursor, and the smaller precursor arises through: 1. post translational cleavage of two N-terminal residues from the larger precursor; 2. initiation of mRNA translation at two points. It is generally thought that Met is the initiator residue in eukaryotes (reviewed in ref. 3). In agreement, we find N-terminal Met in the two precursors of M-41 (this report), and in the precursors of M-321 (3), M-63 (4), and M-104E (unpublished data) L-chains. The proximity of Met-1 to Met-3 permits us to propose that the longer and shorter precursors may originate from initiation of M-41 mRNA translation at the codons specifying Met-1 and Met-3, respectively.

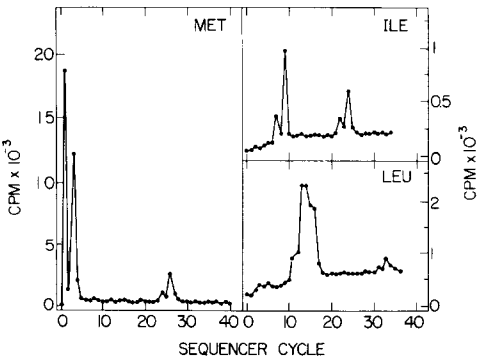

Fig. 4. Radioactivity recovered at each sequencer cycle from the total cell-free products programmed by M-41 mRNA, and labeled with [^{35}S]Met, [^{3}H]Ile, or [^{3}H]Leu.

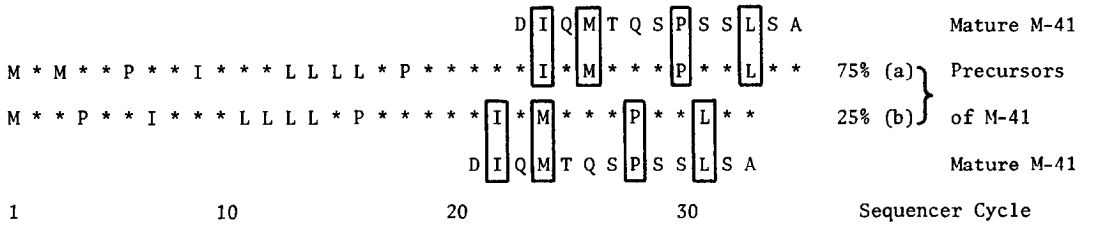

Fig. 5. Amino-terminal sequence of the two precursors of the M-41 L-chain deduced from amino acid sequence analyses of the total cell-free products programmed by the M-41 L-chain mRNA (see Figs. 3,4). Sequence homology between the precursors and the mature L-chain is indicated. The one letter code for amino acids is used (11).

The abundance (20%) and clustering of the Leucine residues in a quadriplet, indicated to us that the extra piece moiety would be quite hydrophobic. Indeed, preliminary results show that at least 65% of the residues in the extra pieces of the M-41 and M-321 L-chain precursors are hydrophobic (Schechter and Burstein, unpublished data). The marked hydrophobicity suggests that the role of the extra piece is to favor interaction of the precursor with the endoplasmic membranes or cell surface (or both).

Sequence Analyses of M-321 L-Chain Precursors. It has been previously shown that the N-terminal extra piece of M-321 L-chain precursor is twenty residues in length, and that it contains one methionine at the N-terminus and six leucines at positions 6, 7, 8, 11, 12, 13 (1-3). For efficient comparison with the data of M-41 precursor, we sequenced the M-321 precursor which was labeled with [^{3}H]Pro and [^{3}H]Ile. The results are given in Fig. 6. The Pro at cycle 28 and the Ile at cycle 22 match with Pro-8 and Ile-2 in the mature M-321 L-chain (5), as expected for an extra piece 20 residues in length. It is also seen that the extra piece of the M-321 precursor contains one proline residue at position 16, and no isoleucine.

Evidence that the Gene Coding for the Variable-Region of Ig L-chains is Larger than Hitherto Known. The identification of extra piece coupled to the V-region of the L-chain immediately raises the question whether it is part of the V-region or whether

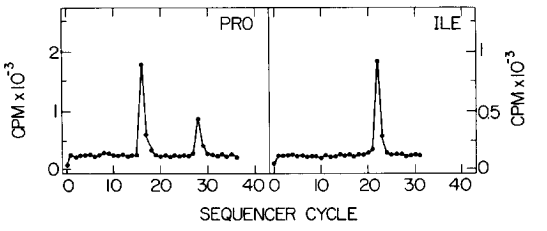

Fig. 6. Radioactivity recovered at each sequencer cycle from the total cell-free products programmed by M-321 mRNA, and labeled with [^{3}H]Pro or [^{3}H]Ile.

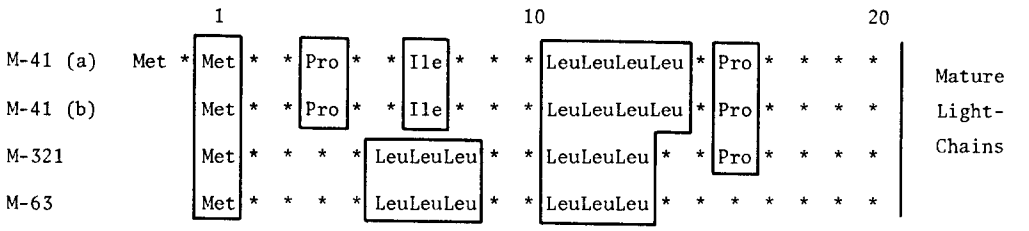

Fig. 7. Comparison of the size and partial sequence of the N-terminal extra piece in the precursors of M-41, M-321 and M-63 L-chains. In M-41 and M-321 the sequence of Met, Leu, Ile and Pro were determined (this report), in M-63 only the sequence of Met and Leu were determined (4). Sequences are lined up to give maximal homology with the N-terminal extra piece of M-321 precursor . Homologous regions are enclosed.

TABLE I. Differences between the precursors of κ-type immunoglobulin light-chains

Precursor	Residues in extra piece	Amino acid substitutions in			
		extra piece (minimal value*)		variable-region+	
		No.	%	No.	%
MOPC-41(a)	22	7/22	32	53/111	48
MOPC-41(b)	20	5/20	25	53/111	48
MOPC-63	20	0	0	8/111	7
MOPC-321	20	0	0	0	0

Amino acid substitutions are compared to M-321.
* Calculated from partial sequences known so far, see Fig. 7.
+ Calculated from amino acid sequences of the M-41, M-63 and M-321 L-chains (5).

it represents another "constant" polypeptide segment. To resolve this issue we determined the size and partial sequence of precursors of L-chains of the same and different subgroups. The M-321, M-63 and M-41 L-chains are of the κ-type and share identical constant-regions (the C-terminal half of the molecule). The M-321 and M-63 are of the same subgroup and differ at the V-region only in 8 out of 111 amino acid residues, the M-41 is of a different subgroup and differs from M-321 V-region in 53 positions (5). Although the sequence data are incomplete (Fig. 7), it is evident that the extra piece of L-chains of different subgroups (M-41 versus M-321 and M-63) differ in size and sequence (Table I), suggesting that the extra piece is part of the V-region. The extra pieces of M-321 and M-63, which are of the same subgroup, are of the same size and so far they share identical Leu and Met sequence. These findings imply that the gene coding for the Ig V-region is larger than hitherto known.

Acknowledgment. The authors thank Ms. Frida Kantor and Ms. Ida Oren for excellent assistance. This research was supported in part by grant 806 from the United States Israel Binational Science Foundation (BSF), Jerusalem, Israel.

REFERENCES

1. Schechter, I. (1973) Proc. Nat. Acad. Sci. 70, 2256-2260.
2. Schechter, I., McKean, D.J., Guyer, R. and Terry, W. (1975) Science 188, 160-162
3. Schechter, I. and Burstein, Y. Biochem. J. In press.
4. Burstein, Y. and Schechter, I. Unpublished data.
5. McKean, D., Potter, M. and Hood, L. (1973) Biochemistry 12, 760-771.
6. Schechter, I. (1974) Biochemistry 13, 1875-1885.
7. Smithies, O., Gibson, D., Fanning, E.M., Goodfliesh, R.M., Gilman, J.G. and Ballantyne, D.L. (1971) Biochemistry 10, 4912-4921.
8. Gray, W.R., Dreyer, W.J. and Hood, L. (1967) Science 155, 465-467.
9. Raftery, M. and Cole, R. (1963) Biochem. Biophys. Res. Commun. 10, 467-472.
10. Maizel, J. (1972) Methods Virol. 5, 179-246.
11. Dayhoff, M. (1972) Atlas of Protein Sequence and Structure, Nat. Biomed. Res. Found., Silver Spring, Md.