Volume 9 Number 18 1981 Nucleic Acids Research

Complete sequence of an immunoglobulin mRNA using specific priming and the dideoxynucleotide method of RNA sequencing

Pamela H.Hamlyn, Michael J.Gait and Cesar Milstein

Laboratory of Molecular Biology, The MRC Centre, Hills Road, Cambridge CB2 2QH, UK

Received 30 June 1981

ABSTRACT

The complete sequence of the mouse immunoglobulin kappa light chain MOPC 21 messenger RNA has been determined using a chain termination method and chemically synthesised deoxyoligonucleotides to initiate the synthesis of a DNA molecule complementary to the mRNA template. Five such oligonucleotide primers have been used for the sequence analysis of this messenger RNA. The approach is excellent for comparative studies of mouse κ-chain mRNAs because they can be made on impure mRNA preparations.

The MOPC 21 light chain mRNA is 943 nucleotides in length excluding the poly(A) region. An unexpected finding was that there are only three bases in the 5' non-coding region and its significance in terms of ribosome binding is discussed; 87 code for the precursor or leader sequence of the protein, 642 for the mature protein and 211 for the 3' non-coding region. The codons for the precursor region allows the previously undetermined amino acid sequence to be predicted. In common with other precursor regions a high proportion of the predicted amino acids are hydrophobic.

INTRODUCTION

Direct analysis of the primary structure of ribonucleic acids is usually performed on material of a high degree of purity. This is easily achieved for viral, ribosomal and some transfer RNAs but is much more difficult for particular messenger RNAs. Since messenger RNAs are often very similar in size their sequence differences must be exploited to make them amenable to sequence analysis. We have adopted a strategy which only requires partially purified mRNA. It consists of priming with an oligonucleotide designed to base-pair solely to the mRNA of interest. The cDNA thus preferentially transcribed is pure enough for limited characterisation [1,2] and sequence analysis by any of the standard methods used for rapid sequence of DNA [3,4]. Using this method of sequence analysis we have established the sequence of constant and 3' non-coding regions of the mRNA for mouse immunoglobulin kappa light chain [5,6]. Other mRNAs have also been partially sequenced using this approach (e.g. ovalbumin [7], globin [8,9]). This report describes the application of the technique of primed synthesis to the V-region of the MOPC 21 light chain mRNA, enabling the first complete primary structure of an immunoglobulin light chain mRNA to be elucidated.

MATERIALS AND METHODS

Materials

Reverse transcriptase, 12,000 u/ml (from avian myeloblastosis virus) was provided by J.W. Beard. ^{32}P-labelled deoxynucleoside triphosphates (400 Ci/ mmole) were from Amersham International Ltd. Deoxynucleoside triphosphates were from Boehringer Chemical Corporation. Dideoxy-nucleoside triphosphates were obtained from P-L Biochemicals.

Preparation of oligonucleotide primers

d(pTAACTGCTCACT) was prepared as described by Gait and Sheppard [10]. d(TGCTCTGGTTT) was prepared as described by Gait *et al*. [11].

Preparation of Ig light chain mRNA

Immunoglobulin light chain mRNA was prepared as described previously [2] except that characterisation of the mRNA by *in vitro* protein synthesis is no longer routine.

Sequencing procedures

The cDNA was sequenced by an adaptation of the chain termination method of DNA sequencing [4] as described previously [6] except that cDNA was synthesised for 15 min at 42°C instead of 30 min at 37°C.

RESULTS

Choice of oligonucleotide primers

In order to determine the nucleotide sequence of an RNA as long as the immunoglobulin mRNA by the method of primed synthesis it is necessary to use several oligonucleotides as primers. The mRNA is sequenced from its 3' end in the direction of the 5' end of the RNA. As the sequence of a region becomes known it is possible to choose a binding site and construct a primer that will allow the sequence to be read farther towards the 5' end of the mRNA. The mRNA for the MOPC 21 light chain was originally estimated to be 1250 (± 100) bases on gel electrophoretic mobility studies [12]. Sequence analysis was initiated from the 3' end using d(T_{10}CA) which gave the whole of the untranslated 3' end, comprising just over 200 residues, and a segment of the constant region. The following 300 residues were sequenced starting with a primer near the end of the translated 3' end of the mRNA corresponding to amino acid residues 213-214 to give the full sequence of the C-region (Fig.

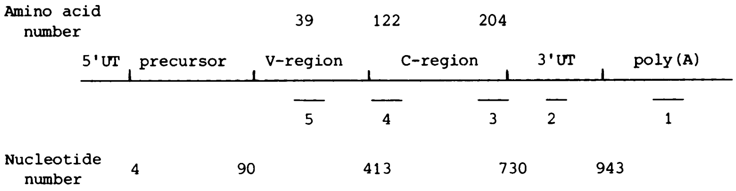

Figure 1. Binding site of oligonucleotides used to prime cDNA synthesis on mouse κ-chain mRNA. The diagram (not to scale) illustrates the binding sites of the five oligonucleotides used to prime the synthesis of cDNA used for sequence analysis. The oligonucleotides are: 1) d(T_{10}CA), 2) d(pGGAGGAGAA), 3) d(TTGGGT), 4) d(pTAACTGCTCACT), 5) d(TGCTCTGGTTT).

1). The remainder of the mRNA includes 400 ± 50 nucleotides comprising the V-region, the precursor region and the 5' untranslated region, thought to be "quite short" [5]. Previous experience of sequence analysis using a heterogeneous template (e.g. an impure preparation of light chain mRNA) and a short primer had indicated that only about 300 bases, at most, could be accurately determined from one priming site. It was therefore necessary to use two different oligonucleotides as primers in order to elucidate the entire sequence (see Fig. 1). The two priming sites were chosen as follows. From the nucleotide sequence of the constant region the primer d(pTAACTGCTCA CT) was synthesised and shown to initiate transcription in the constant region corresponding to amino acid residues 122-125, about 14 amino acids from the V/C junction. Its priming site was chosen so that a gel reading would include some of the constant region and thus verify the binding site. More important, being in the constant region but transcribing the variable region would allow the primer to provide a method of comparative analysis of all mouse kappa light chain variable regions.

The other oligonucleotide which was synthesised, d(TGCTCTGGTT), primes in the V-region. It was designed to bind to the mRNA where it codes for a framework region of the protein (amino acids 39-42). This oligonucleotide has been tested with a kappa light chain mRNA of unknown V-region sequence and it does appear to prime in the V-region at the same position as in MOPC 21 variable region but not as specifically as with MOPC 21 (results not shown. In the present study the primers have been used for sequence analysis by the dideoxynucleotide inhibition method. In this way the nucleotides next to the primer (about 20 on the 3' extension of the primer) are usually not determined. If the sequence read from the gels at the primer plus 20 nucleotides position is the one expected (as deduced from the amino acid sequence) then it is assumed that the primer is binding in the predicted position.

Sequence analysis using d(pTAACTGCTCACT) and d(TGCTCGGTTT)

Each primer is used separately to initiate cDNA transcription in four different sequencing reactions containing inhibitors of chain elongation specific for each of the four bases. The oligonucleotide d(pTAACTGCTCACT) base pairs to the mRNA coding for amino acid residues 122-125 and allows the sequence of the V/C boundary to be determined and up to residue 26. The other oligonucleotide primes at residues 39-42 and allows the sequence to be determined from about amino acid 32 to the 5' terminus of the mRNA. In other words, the region amino acid 26-32 was sequences by both primers. The complete sequence is shown in Figure 2. In the study of the V-region some nucleotides were difficult to ascertain. This was because there was a band in each of the four nucleotide lanes so that the correct nucleotide could not be decided. If the sequencing reactions are made with two different radioactive nucleotides and the two series run in parallel these ambiguities do sometimes occur in different places so that where, for example, a particular nucleotide is obscured in a C-labelled reaction it may be easily determined in an A-labelled reaction (and *vice versa*). Even so, there remained one base corresponding to the Ser-34 and another at Pro-59 which could not be determined.

The sequence analysis of the V-region confirms the amino acid sequence determined by Svasti and Milstein [13]. Immunoglobulin light chain is first synthesised with a precursor which is cleaved off during processing of the protein and is therefore not present in the secreted product [14,15]. To determine the amino acid composition of this precursor region requires that the mRNA be translated *in vitro* in a cell-free system which does not contain the appropriate cleavage enzymes. The radiolabelled protein is then sequenced from its N-terminus. Although this has been achieved for other immunoglobulin light chains [16] the immunoglobulin light chain from MOPC 21 has not been sequenced although its total length has been determined (29 amino acid residues) as well as suggested positions for three methionines at residues -29, -24 and -20 [17]. When the nucleotide sequence data obtained using the primer d(TGCTCTGGTTT) is translated into protein it is in complete agreement with these observations. As in other examples, the precursor sequence has a high proportion of hydrophobic amino acids as predicted for the leader sequence of secreted proteins. Early work with the mRNA for the MOPC 21 light chain was done on ^{32}P internally labelling mRNA and the now practically abandoned T1 ribonuclease digestion in fingerprint analysis [12].

```
      -29                                                   -20
      Met   His   Gln   Thr   Ser   Met   Gly   Ile   Lys   Met   Glu   Ser   Gln   Thr   Leu   Val   Phe   Ile   Ser
N N N A U G C A U C A G A C C A G C A U G G G C A U C A A G A U G G A A U C A C A G A C U C U G G U C U U C A U A U C C
                  10                  20                  30                  40                  50                  60
-10                                                         1
Ile   Leu   Leu   Trp   Leu   Tyr   Gly   Ala   Asp   Gly   Asn   Ile   Val   Met   Thr   Gln   Ser   Pro   Lys   Ser
A U A C U G C U C U G C U U A U A U G G A G C U G A U G G G A A C A U U G U A A U G A C C C A A U C U C C C A A A U C C
                  70                  80                  90                 100                 110                 120
11                                                          21
Met   Ser   Met   Ser   Val   Gly   Glu   Arg   Val   Thr   Leu   Thr   Cys   Lys   Ala   Ser   Glu   Asn   Val   Val
A U G U C C A U G U C A G U A G G A G A G A G G G U C A C C U U G A C C U G C A A G G C C A G U G A G A A U G U G G U U
                 130                 140                 150                 160                 170                 180
31                                                          41
Thr   Tyr   Val   Ser   Trp   Tyr   Gln   Gln   Lys   Pro   Glu   Gln   Ser   Pro   Lys   Leu   Leu   Ile   Tyr   Gly
A C U U A U G U U U C N U G G U A U C A A C A G A A A C C A G A G C A G U C U C C U A A A C U G C U G A U A U A U G G G
                 190                 200                 210                 220                 230                 240
51                                                          61
Ala   Ser   Asn   Arg   Tyr   Thr   Gly   Val   Pro   Asp   Arg   Phe   Thr   Gly   Ser   Gly   Ser   Ala   Thr   Asp
G C A U C C A A C C G G U A C A C U G G G G U C C C N G A U C G C U U C A C A G G C A G U G G A U C U G C A A C A G A U
                 250                 260                 270                 280                 290                 300
71                                                          81
Phe   Thr   Leu   Thr   Ile   Ser   Ser   Val   Gln   Ala   Glu   Asp   Leu   Ala   Asp   Tyr   His   Cys   Gly   Gln
U U C A C U C U G A C C A U C A G C A G U G U G C A G G C U G A A G A C C U U G C A G A U U A U C A C U G U G G A C A G
                 310                 320                 330                 340                 350                 360
91                  V ←─┼─→ J                               101                             J ←─┼─→ C
Gly   Tyr   Ser   Tyr   Pro   Tyr   Thr   Phe   Gly   Gly   Gly   Thr   Lys   Leu   Glu   Ile   Lys   Arg   Ala   Asp
G G U U A C A G C U A U C C G U A C A C G U U C G G A G G G G G C A C C A A G C U G G A A A U A A A A C G G G C U G A U
                 370                 380                 390                 400                 410                 420
111                                                         121
Ala   Ala   Pro   Thr   Val   Ser   Ile   Phe   Pro   Pro   Ser   Ser   Glu   Gln   Leu   Thr   Ser   Gly   Gly   Ala
G C U G C A C C A A C U G U A U C C A U C U U C C C A C C A U C C A G U G A G C A G U U A A C A U C U G G A G G U G C C
                 430                 440                 450                 460                 470                 480
131                                                         141
Ser   Val   Val   Cys   Phe   Leu   Asn   Asn   Phe   Tyr   Pro   Lys   Asp   Ile   Asn   Val   Lys   Trp   Lys   Ile
U C A G U C G U G U G C U U C U U G A A C A A C U U C U A C C C C A A A G A C A U C A A U G U C A A G U G G A A G A U U
                 490                 500                 510                 520                 530                 540
151                                                         161
Asp   Gly   Ser   Glu   Arg   Gln   Asn   Gly   Val   Leu   Asn   Ser   Trp   Thr   Asp   Gln   Asp   Ser   Lys   Asp
G A U G G C A G U G A A C G A C A A A A U G G C G U C C U G A A C A G U U G G A C U G A U C A G G A C A G C A A A G A C
                 550                 560                 570                 580                 590                 600
171                                                         181
Ser   Thr   Tyr   Ser   Met   Ser   Ser   Thr   Leu   Thr   Leu   Thr   Lys   Asp   Glu   Tyr   Glu   Arg   His   Asn
A G C A C C U A C A G C A U G A G C A G C A C C C U C A C G U U G A C C A A G G A C G A G U A U G A A C G A C A U A A C
                 610                 620                 630                 640                 650                 660
191                                                         201
Ser   Tyr   Thr   Cys   Glu   Ala   Thr   His   Lys   Thr   Ser   Thr   Ser   Pro   Ile   Val   Lys   Ser   Phe   Asn
A G C U A U A C C U G U G A G G C C A C U C A C A A G A C A U C A A C U U C A C C C A U U G U C A A G A G C U U C A A C
                 670                 680                 690                 700                 710                 720
211
Arg   Asn   Glu   Cys   term.
A G G A A U G A G U G U U A G A G A C A A A G G U C C U G A G A C G C C A C C A C C A G C U C C C C A G C T C C A T C C
                 730                 740                 750                 760                 770                 780

U A U C U U C C C U U C U A A G G U C U U G G A G G C U U C C C C A C A A G C G A C C U A C C A C U G U U G C G G U G C
                 790                 800                 810                 820                 830                 840

U C C A A A C C U C C U C C C C A C C U C C U U C U C C U C C U C C U C C C U U U C C U U G G C U U U U A U C A U G C U
                 850                 860                 870                 880                 890                 900

A A U A U U U G C A G A A A A U A U U C A A U A A A G U G A G U C U U U G C A C U U G Poly(A)
                 910                 920                 930                 940
```

Figure 2. The complete nucleotide sequence of mouse immunoglobulin κ-chain mRNA. Nucleotides 412-943 have been published previously [6]. In the sequence 1-412 five nucleotides could not be determined. These are denoted by N and underlined.

A number of oligonucleotides were analysed at that time down to octanucleotides. The correspondence between those results and the ones described in this paper is complete except for one single omission in the old analysis of a theoretical T1 digestion product which should have been detected but was not. The nucleotide sequence of the precursor predicts an RNase T1 product (UCUUCAUAUUCAUACUG) and this was the only large T1 oligonucleotide not detected in the original analysis. A sequence error cannot be totally excluded since it is based on a single method of analysis but we feel that the difference is probably due to two factors in the original analysis. One was that it was very near the 5' end of the mRNA and therefore nicked from a fraction of the mRNA. The other factor is the unusually large number of U residues in the sequence rendering it particularly susceptible to pancreatic RNase.

The 5' untranslated region

It is assumed that the end of the cDNA represents the 5' end of the mRNA. This end point is deemed to be reached where there is a strong band in each slot in the gel. At this point the chain termination method breaks down and three residues cannot be determined (Fig. 3). From the results of Figure 3 and other gels we concluded that there are three nucleotides to the 5' end of the AUG initiation codon but their identity remains unknown. The fainter bands which run behind the strong stop band may be due to heterogeneity of the 5' end of the RNA. A similar problem was encountered by McReynolds et al. [18] when using the same method to determine the 5' end of ovalbumin mRNA. This mRNA has recently been shown to be heterogeneous at its 5' end [19]. In our case the difficulty is exacerbated by the greater heterogeneity of the template and the shorter length (and therefore, presumably decreased specificity) of the oligonucleotide primer.

In view of the similarity of the methods and results with the above mentioned authors we conclude that, as in ovalbumin, the light chain mRNA starts with a cap structure (which, as in most mRNAs, is likely to be $7^{m}G$)

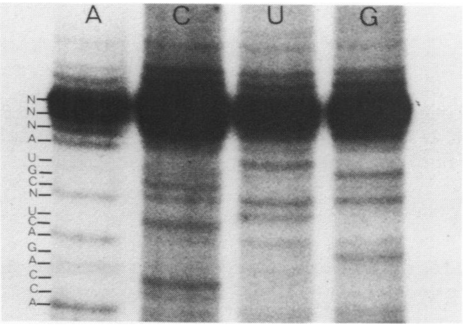

Figure 3. Autoradiograph of termination of cDNA transcription using d(TGCTCTGGTTT). Using the primer d(TGCTCTGGTTT) a sequence was generated and displayed on an autoradiograph of a gel. The sequence determined (written as RNA) was: N N N A U G C N U C A G A C C A. (From other experiments the
          1   10
nucleotide at position 8 was found to be A.) The strong band at position 1 is taken to be the termination of transcription of the cDNA and thus the 5' end of the mRNA.

which precedes the beginning of the cDNA. For this reason our final structure is written as shown in Figure 2, namely (^{m}G) NNNAUG ...

DISCUSSION

Methodological aspects

Using the dideoxy method of nucleotide sequence analysis it has been possible to determine the sequence of nucleotides in the 943 [excluding poly(A)] long immunoglobulin light chain mRNA. During this study, five oligonucleotide primers have been used. With the experience gained we consider that only four are necessary to completely sequence any new kappa chain mRNA. A more widely used method of sequence analysis of mRNAs is to prepare a double-stranded DNA, one strand being complementary to the mRNA, and form a recombinant DNA molecule with some vector, then clone it. This can provide completely adequate information on the coding region of the molecule but leads to difficulties in determining the sequence at the 3' and 5' ends of the molecule. For instance, ovalbumin mRNA, although mainly sequenced using cloned cDNA, had the 3' and 5' ends sequenced using specific primer transcription and other methods [18]. Similarly, to verify the bases adjacent to the poly(A) tail of an immunoglobulin heavy chain mRNA another sequencing method had to be used other than sequence analysis of the cloned cDNA [20]. Most important, sequence analysis of new light chains on which the same primers could be used is a much simpler and faster approach for comparative purposes.

As far as design of new primers, apart from the usual considerations of sequence, namely the need to avoid the ones which make chemical synthesis difficult and primers poorly soluble or self complementary, the major consideration is sufficient base pairing to give specificity and efficiency of priming. The longer the primer the more likely this will be achieved, but specific cases of high efficiency are possible with some short sequences, e.g. the priming in the C region (3 in Fig. 1). These are largely dependent on favourable secondary structure and since this is unknown, short primers are generally unsuccessful and troublesome because they prime at sites other than the ones they were designed for. This is a major consideration to be kept in mind and proper identification of priming site is essential, before primers of lengths up to at least 15-20 bases long can be relied upon for further work.

The primed synthesis method of sequence analysis provides no information on modified bases neither a method of cross checking the sequence. Corroboration is obtained by amino acid analysis and previous classical methods of nucleic acid sequence determination; but this is incomplete. A possible method of independently checking the sequences is to make full length cDNA and isolate it as a unique band from a polyacrylamide gel. If the cDNA has been previously radiolabelled at its 5' end to a high specificity activity it can be sequenced by the chemical degration method of Maxam and Gilbert [3]. A major drawback of this method is the large quantity of RNA required in the initial cDNA transcription reaction [21].

Significance of sequence

The most unexpected feature of the sequence was the shortness of the untranslated 5' end. The evidence presented here on this point is unfortunately not conclusive. The possibility that the strong band is due to a secondary structure effect or to specific degradation of the mRNA is not excluded. However, we find this unlikely based on the experience which we and others have in using the method. Further evidence will require studies of cap sequences, comparative studies with other light chains, mRNA etc. There is one example in a Sindbis virus protein where an AUG follows the cap structure, but this AUG is not the initiation point of translation [22]. In the case of the mRNA for a vesicular stomatitis viral protein, the initiator AUG is preceded by nine residues of untranslated 5' end [23]. Therefore the four untranslated bases of the light chain (including the presumed cap) constitute the shortest untranslated 5' sequence so far. That translation starts in the first AUG, as shown in Figure 2, is based on the positioning of methionines in translated mRNA in an *in vitro* system. Rose *et al.* [17] have analysed by automatic sequencing methods, the positions at which radioactive methionine was incorporated. The results at various positions were not compatible with the presence of a single polypeptide. Indeed the primary transcript of MOPC 21 mRNA containing the precursor to the light chain was a closely spaced doublet on polyacrylamide gel electrophoresis [14]. The results were therefore compatible with either initiation at two sites or partial degradation and Rose *et al.* [17] concluded that the methionines were at positions -29, -24 and -20. This is fully confirmed by the results of this paper. The origin of the heterogeneity remains obscure. Whether the presence of a very short 5' end has any bearing on this matter is doubtful. But the shortness of the 5' end and the fact that initiation on other nearby in-phase AUGs does not seem to take place, provides further evidence for the suggestion that the AUG itself is an important component of the ribosome binding site and that initiation is on the first available AUG [24,25].

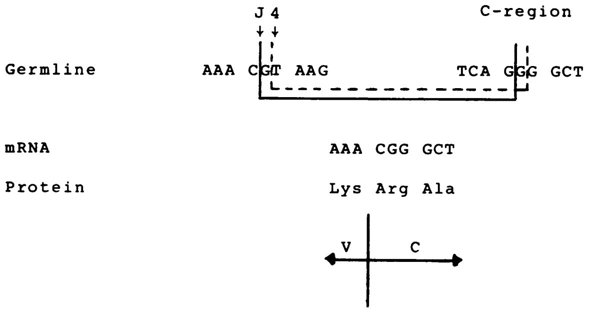

Figure 4.

It is interesting to compare the precursor sequence of MOPC 21 with others, particularly as far as the distribution of methionines is concerned. Met at position -20 (the third in MOPC 21) is present in all the sequences of precursors listed [16] and is the initiator methionine in 11 out of 15 examples. In one example, initiation is on the -24 Met, also found in MOPC 21 but not as initiator. The immunoglobulin light chain from MPC 11 starts at the same -29 Met as MOPC 21 and also shows N-terminal heterogeneity [17]. It will be interesting to compare the mRNA sequences in other cases. This may throw some further light on the requirements for initiation of translation.

Concerning the rest of the sequence, the nucleotide sequence between amino acid residues 96 and 107 corresponds exactly with the J4 germline sequence as described by Max et al. [26,27]. (This J-region is called J2 by Sakano et al. [28].) Arg-107 is the residue usually recognised as the first of the C-region gene and the mRNA sequence demonstrates that at least the third base of the Arg must come from the C-region DNA. The second base could be donated by either the J-gene or the C-region gene, as shown in Figure 4.

Therefore, splicing must have occurred at one of the alternative pairs indicated with arrows, both of which are untypical, although the one shown with a continuous line is more likely in view of its comparison with the consensus splice sequences taken from Lewin [29]. The subscripts indicate the percent occurrence of the most common base:-

$$A_{46}\ A_{80}\ G_{90}\overset{\downarrow}{\ }G_{100}\ T_{100}\ \cdots\cdots\cdots\ C_{67}\ A_{97}\ G_{100}\overset{\downarrow}{\ }G_{49}$$

Note that the G (90% frequency) preceding the splice point in MOPC 21 is C.

REFERENCES

1. Cheng, C.C. Brownlee, Carey, N., Doel, M.T., Gillam, S. and Smith, M. (1976) J. Mol. Biol. 107, 525-547.
2. Hamlyn, P.H., Gillam, S., Smith, M. and Milstein, C. (1977) Nucleic Acids Res. 4, 1123-1134.
3. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
4. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
5. Milstein, C., Chen, K.C.S., Hamlyn, P.H., Rabbitts, T.H. and Brownlee, G.G. (1977) ICN-UCLA Symp. on Molecular and Cellular Biology, Vol. VI, pp. 29-41, Academic Press, New York.
6. Hamlyn, P.H., Brownlee, G.G., Cheng, C.C., Gait, M.J. and Milstein, C. (1978) Cell 15, 1067-1075.
7. Brownlee, G.G. and Cartwright, E.M. (1977) J. Mol. Biol. 114, 93-117.
8. Baralle, F.E. (1977) Cell 12, 1085-1095.
9. Proudfoot, N.J. (1977) Cell 10, 559-570.
10. Gait, M.J. and Sheppard, R.C. (1979) Nucleic Acids Res. 6, 1259-1268.
11. Gait, M.J., Singh, M., Sheppard, R.C., Edge, M.D., Greene, A.R., Heathcliffe, G.R., Atkinson, T.C., Newton, C.R. and Markham, A.F. (1980) Nucleic Acids Res. 8, 1081-1096.
12. Milstein, C., Brownlee, G.G., Cartwright, E.M., Jarvis, J.M. and Proudfoot, N.J. (1974) Nature 252, 354-359.
13. Svasti, J. and Milstein, C. (1972) Biochem. J. 128, 427-444.
14. Milstein, C., Brownlee, G.G., Harrison, T.M. and Mathews, M.B. (1972) nature 239, 117-120.
15. Harrison, T.M., Brownlee, G.G. and Milstein, C. (1974) Eur. J. Biochem. 47, 613-620.
16. Kabat, E.A., Wu, T.T. and Bilofsky, H. (1979) in "Sequences of Immunoglobulin genes", N.I.H. Publication No. 80-2008.
17. Rose, S.M., Kuehl, W.M. and Smith, G.T. (1977) Cell 12, 453-463.
18. McReynolds, L., O'Malley, W.B., Nisbet, A.D., Fothergill, J.C., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723-728.
19. Malek, L.T., Eschenfeldt, W.H., Munns, T.W. and Rhoads, R.E. (1981) Nucleic Acids Res. 9, 1657-1673.
20. Dunnick, W., Rabbitts, T.H. and Milstein, C. (1980) Nucleic Acids Res. 8, 1475-1484.
21. Noyes, B.E., Mevarech, M., Stein, R. and Agarwal, K.K. (1979) Proc. Nat. Acad. Sci. USA 76, 1770-1774.
22. Hefti, E., Bishop, D.H.L., Dubin, D.T. and Stollar, V. (1976) J. Virol. 17, 145-159.
23. Rose, J.K. (1978) Cell 14, 345-353.
24. Baralle, F.E. and Brownlee, G.G. (1978) Nature 274, 84-87.
25. Kozak, M. (1978) Cell 15, 1109-1123.
26. Max, E.E., Seidman, J.G. and Leder, P. (1979) Proc. Nat. Acad. Sci. USA 76, 3450-3454.
27. Max, E.E., Seidman, J.G., Millar, H. and Leder, P. (1980) Cell 21, 793-799.
28. Sakano, H., Hüppi, K., Heinrich, G. and Tonegawa, S. (1979) Nature 280, 288-294.
29. Lewin, B. (1980) Cell 22, 324-326.