

of the nervous system²⁴. As NGF reversibly increases cell-substratum adhesion, growing axons could be 'guided' by the NGF-induced increase in the adhesiveness of the growth cone to the underlying cells, rather than the growth cone following a substratum-associated gradient. Growth cones preferentially elongate on surfaces to which they adhere more firmly²⁵, and it should make no difference if the growth cone or substratum is responsible for initiating the adhesive response. This type of NGF effect on cell-substratum adhesiveness could also explain the requirement for local exposure to NGF for continued neurite outgrowth in cultured sympathetic ganglion cells²⁶.

Finally, a similarity between three groups of clonal cell lines which are developmentally responsive to external stimuli should be noted. The cells are PC12, C1300 mouse neuroblastoma and the Friend erythroleukaemic lines. As outlined above, NGF, cyclic AMP and compounds which induce calcium mobilisation in PC12 cells lead to developmental changes. In C1300 neuroblastoma cells there is a well defined relationship between cell-substratum adhesiveness and morphological differentiation⁹ and such compounds as cyclic AMP, butyric acid, dimethyl sulphoxide (DMSO), prostaglandin E₁, acetylated diamines and ouabain (D.S., unpublished observation) also induce neurite extension^{9,27}. A set of reagents overlapping the above also induce erythroid differentiation in Friend virus-transformed erythroleukaemic cells. These include DMSO, short-chain fatty acids, polymethylene acetamide, prostaglandin E₁, and ouabain (see ref. 28 for discussion). Although it is extremely difficult to envisage a direct action of all these diverse conditions and reagents on the genome of the cells, it is likely that all interact to varying extents with the plasma membrane. It has, for example, been demonstrated that DMSO, divalent ions, and other cryoprotective agents produced an increase in the phase transition temperature of membranes which correlated with their inductive effects on erythroid differentiation²⁹. Thus, it is likely that all of these reagents produce their phenotypic effects by initially altering

the membrane structure, perhaps by causing a change in calcium ion permeability. The requisite next step would be the flow of this information acquired by the membrane to the biosynthetic machinery of the cell. The mechanism of this transduction remains to be described.

β -NGF and its antiserum were obtained from R. Goodman and H. Herschman. Supported by the Muscular Dystrophy Association.

Received 9 December 1977; accepted 24 April 1978.

1. Levi-Montalcini, R. in *Golgi Centennial Symposium* (ed. M. Santini) 437-448 (Raven, New York, 1975).
2. Greene, L. A. *Devl Biol.* **58**, 96-105 (1977).
3. Greene, L. A. & Tischler, A. S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2424-2428 (1976).
4. Schubert, D., Heinemann, S. & Kidokoro, Y. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2579-2583 (1977).
5. Greene, L. A. & Rein, G. *Nature* **268**, 351-353 (1977).
6. Schubert, D. & Klier, F. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5184-5188 (1977).
7. Mains, R. E. & Patterson, P. H. *J. Cell. Biol.* **59**, 329-345 (1973).
8. Johnson, M. et al. *Nature* **262**, 308-310 (1976).
9. Patrick, J., Heinemann, S. & Schubert, D. A. *Rev. Neurosci.* **1**, 417-443 (1978).
10. Schubert, D. & Whitlock, C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4055-4058 (1977).
11. Messer, A. *Brain Res.* **130**, 1-12 (1977).
12. Li, A. P., O'Neill, P., Kawashima, K. & Hsie, A. W. *Archs Biochem. Biophys.* **182**, 181-187 (1977).
13. Stallcup, W. J. *Physiol., Lond.* (submitted).
14. Schudt, C., Gaertner, U. & Pette, D. *Eur. J. Biochem.* **68**, 103-111 (1976).
15. Obrink, B., Kuhlenschmidt, M. S. & Roseman, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1077-1081 (1977).
16. Garrels, J. E. & Schubert, D. *J. biol. Chem.* (submitted).
17. Furmanski, P., Silberman, D. J. & Lubin, M. *Nature* **233**, 413-414 (1971).
18. Schubert, D. *Neurobiology* **4**, 376-387 (1974).
19. Shields, R. & Pollock, K. *Cell* **3**, 31-38 (1974).
20. Gingle, A. R. *Devl Biol.* **58**, 394-401 (1977).
21. Rasmussen, H. & Goodman, D. *Physiol. Rev.* **57**, 421-516 (1977).
22. Anderson, D. R., Davis, J. L. & Caraway, K. L. *J. Biochem.* **252**, 6617-6623 (1977).
23. Eisen, H., Nasi, S., Georgopoulos, C. P., Arndt-Jovin, D. & Ostertag, W. *Cell* **10**, 689-695 (1977).
24. Levi-Montalcini, R. *Progress in Brain Res.* Vol. 45 (eds Corner, M. A. & Swaab, D. F.) 235-258 (Elsevier, Amsterdam, 1976).
25. Letourneau, P. C. *Devl Biol.* **44**, 92-101 (1975).
26. Campenot, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4516-4519 (1977).
27. Palfrey, C., Kimhi, Y. & Littenhauer, U. Z. *Biochem. biophys. Res. Commun.* **76**, 937-945 (1977).
28. Tabuse, Y., Furusawa, M., Eisen, H. & Shibata, K. *Expl Cell Res.* **108**, 41-45 (1977).
29. Lyman, G. D., Preisler, H. D. & Paphadjopoulos, D. *Nature* **262**, 360-363 (1976).
30. McFarlane, A. S. *Nature* **182**, 53 (1958).

Sequence of chicken ovalbumin mRNA

L. McReynolds*, B. W. O'Malley*, A. D. Nisbet†, J. E. Fothergill†, D. Givol‡, S. Fields§, M. Robertson§ & G. G. Brownlee§

*Baylor College of Medicine, Houston, Texas

†Department of Biochemistry, University of Aberdeen, Aberdeen, UK

‡The Weizmann Institute, Rehovoth, Israel

§MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK

The complete sequence of chicken ovalbumin mRNA is presented; it is 1,859 residues long, excluding its terminal 'cap' and poly(A). The region coding for ovalbumin lies close to the 'cap' but is separated from the poly(A) by an extensive 3' noncoding region of 637 nucleotides which may have no function that is precisely dependent on its sequence.

CHICKEN ovalbumin and its gene provide a favourable and interesting model system for studying the regulation of gene action by hormones^{1,2}. It is firmly established³ that steroid hormones will switch on the expression of all the egg proteins, including ovalbumin, and that the hormone probably acts at the DNA level through a steroid receptor protein. However, the detailed molecular mechanism and details of the specific site or sites involved on the DNA are unknown.

As a first approach to understanding the basic molecular biology of this gene we decided to sequence the messenger RNA. Since we started this work it has become clear that the ovalbumin gene⁴⁻⁷, as well as other genes such as the β -globin^{8,9} and immunoglobulin genes¹⁰ are interrupted by 'introns'¹¹ or inserted sequences. This complication was unexpected and made it essential that an accurate sequence of the mRNA be established for future comparison with gene sequences.

Previous sequencing of ovalbumin mRNA has been limited to 130 residues immediately adjacent to the poly(A), with a further tentative sequence from 131 to 191 residues from the poly(A)^{12,13}. This study revealed the presence of the sequence A-A-U-A-A-A centred about 20 residues away from the poly(A)¹⁴, although the function of this signal remains unknown.

The sequence work described here was possible because of the synthesis and isolation of an almost full length duplex DNA copy of the mRNA¹⁵ and its subsequent cloning in the

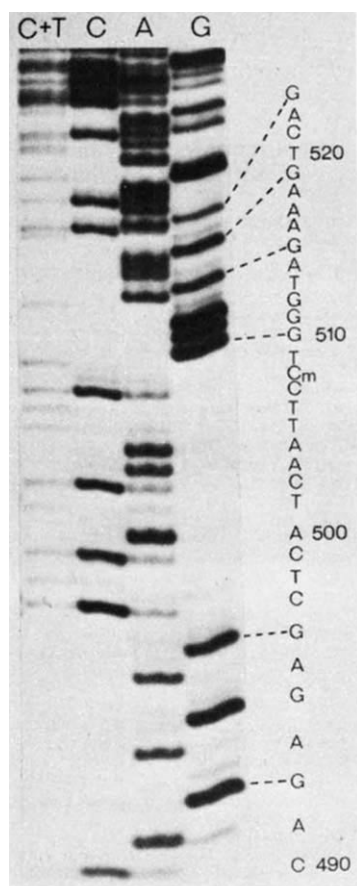


Fig. 1 Sequence derived from residues 490–523 from *Pst*-labelled *Hae*III-recut 341-long fragment. G and A refer to the 'alternate G' and 'alternate A' procedures, whilst C and C+T are the standard hydrazine procedures¹⁸. Residue 508 is methylated as it is part of an *Eco*RII recognition site, and the resultant 5-methyl C (Cm) appears as a gap, or very weak C. 40 µg of pOv230 was opened as its sole *Pst*I site by restriction with 60 units *Pst*I (New England Biolabs) for 1 h at 37 °C in a volume of 200 µl. An equal volume of phenol was added and, after vortex mixing and centrifugation to separate the two layers, the linear DNA was precipitated from the aqueous layer by the addition of one-tenth volume of 3 M sodium acetate and 3 volumes of ethanol. The DNA was treated with excess bacterial alkaline phosphatase (Sigma) (20 µg) in the presence of venom phosphodiesterase (0.5 µg) in 0.01 M Tris-acetate 0.01 M MgCl₂, pH 8.5 in a 100 µl reaction for 30 min at 37 °C. The enzymes were removed and the DNA precipitated by a similar phenol inactivation to that described above for the restriction enzyme. Labelling was carried out using 200 µCi γ-³²P-ATP (3,000–5,000 Ci mmol⁻¹, Radiochemical Centre, Amersham) in a volume of 25 µl using 5–20 units of T4-polynucleotide kinase (P.L. Biochemicals or Boehringer) in 1×glycine-NaOH buffer, pH 9.5 (2×buffer is 100 mM glycine-NaOH, pH 9.5, 20 mM MgCl₂, 10 mM dithiothreitol, 0.2 mM spermidine, 50% glycerol) for 30 min at 37 °C. The total reaction mixture, after the addition of 20 µl H₂O and 0.5 µl of 10 mM ATP, was treated with phenol to remove enzyme and the labelled DNA precipitated, as before, in the presence of 0.3 M ammonium acetate. Recutting was with 100 units of *Hae*III (New England Biolabs) for 1 h at 37 °C in a volume of 20 µl, after which 5 µl of a 30% sucrose solution containing marker dyes (bromophenol blue and xylene cyanol FF) was added. The sample was loaded onto a 3 mm×1.5 mm well of a nondenaturing 4% and 40 cm long acrylamide gel in Tris-borate-EDTA buffer¹⁸ and run for 6 h at 400 kV. The faster band running close to the xylene cyanol FF (slow) dye was eluted¹⁸ and subjected to the Maxam and Gilbert procedure and fractionation on a 20% acrylamide gel. Sensitive autoradiography with phosphotungstate screens at –80 °C was used to detect bands⁴⁰.

Escherichia coli plasmid pMB9¹⁶ using the poly(A)/poly(T) tailing method. Plasmids were screened for the size of inserted sequence and the longest clone (named pOv230) containing about 1,850 residues of ovalbumin mRNA was selected for sequencing. Similar plasmids have been isolated by other workers¹⁷.

Ovalbumin sequences in clone pOv230

Clone pOv230 was sequenced by the chemical degradation method of Maxam and Gilbert¹⁸. An initial difficulty was that there was no satisfactory method of excising the inserted ovalbumin sequence from the vector plasmid. This problem was overcome in one approach by the isolation in bulk of two large *Hae*III restriction fragments which contained mainly ovalbumin-derived sequences. Another approach was to radioactively label a restriction enzyme digest of the entire pOv230 and then purify the specific ovalbumin-containing fragments. Full details of the sequence analysis will be published elsewhere, but as an example, the sequence established by labelling the single *Pst*I site in the intact pOv230 is shown in Fig. 1. Similar labelling of restriction fragments by six different restriction enzymes known to cut ovalbumin sequences¹⁵, that is, *Hae*III, *Hinf*I, *Alu*I, *Mbo*II, *Hph*I and *Hga*I, allowed the entire ovalbumin-derived part of the plasmid to be sequenced (residues 13–1,859 of Fig. 2). A computer search¹⁹ of this sequence showed the final positions (see Table 1) of cutting sites, not only for those enzymes used in the sequence analysis, but also of all other presently known restriction enzymes. Considering the large number of sites, there was fair agreement with the previously published restriction map of ovalbumin cDNA¹⁵ although additional *Xba*I, *Pvu*II and *Taq*I sites were established and more sites were present for some of the enzymes than was previously thought.

Sequence accuracy

The sequence of the coding region of the mRNA (residues 65–1,222) was considerably helped by a comparison of the nucleotide sequence with known amino acid sequence. Several selected regions of amino acid sequence such as those of the N

Table 1 Known restriction recognition sites in pOv 230 (residues 13–1,859 of Fig. 1)*

Enzyme	Residue nos†	Sequence
<i>Pst</i> I	477	CTGCA/G
<i>Hae</i> III	818	GG/CC
<i>Hinf</i> I	258, 566, 1458, 1732	G/ANTC
<i>Hga</i> I	1125	GCGTC
<i>Hph</i> I	60, 336, 1605	TCACC
<i>Alu</i> I	15, 38, 117, 175, 277, 475, 495, 557, 759, 828, 986, 1042, 1235, 1332, 1338, 1554	AG/CT
<i>Mbo</i> II	147, 308, 389, 603, 638, 748, 887, 895, 922, 989, 1030, 1133, 1161, 1197, 1229, 1669	GAAGA or TCTTC
<i>Eco</i> RII‡	193, 213, 254, 507, 754	/CC(A/T)GG
<i>Mbo</i> I‡	482, 751, 898	/GATC
<i>Taq</i> I	41	T/CGA
<i>Xba</i> I	1345	T/CTAGA
<i>Sst</i> I	116, 494	GAGCT/C§
<i>Pvu</i> II	474	CAG/CTG
<i>Asu</i> I	867	G/GACC
<i>Mnl</i> I	325, 444, 447, 892, 914, 929, 943, 1014, 1085, 1103, 1159	CCTC or GAGG

* No sites for *Hae*II, *Hind*II and III, *Hha*I, *Hpa*I and II, *Eco*RI, *Ava*I, *Bam*I, *Bal*I, *Bgl*II, *Sma*I, *Xho*I and *Sal*I.

† The number given is the first (5') nucleotide of the recognition sequence and not the point of cutting. Where this is within the recognition sequence it is indicated by a vertical bar.

‡ These enzymes do not cut in pOv230 grown in *E. coli* strain RR1⁴² because of methylation of the C (*Eco*RII) or the A (*Mbo*I) residues.

§ Recognition sequence established by C. D. Liarakos (personal communication). For others see ref. 43.

Fig. 2 Sequence of ovalbumin messenger RNA. Residues 13–1,859 were present in the plasmid pOv230; residues 1–12 and 1,731–1,859 were sequenced directly from the mRNA. * Marks the terminator, P the serine phosphate residues, and 'CHO' the carbohydrate. Mature ovalbumin lacks the N-terminal methionine; the resultant N-terminal glycine is N-acetylated. The amino acid composition, excluding the initiator Met, is Ala (A), 35; Arg (R), 15; Asn (N), 17; Asp (D), 14; Cys (C), 6; Gln (Q), 15; Glu (E), 33; Gly (G), 19; His (H), 7; Ile (I), 25; Leu (L), 32; Lys (K), 20; Met (M), 16; Phe (F), 20; Pro (P), 14; Ser (S), 38; Thr (T), 15; Trp (W), 3; Tyr (Y), 10 and Val (V), 31. Total, 385 residues.

mined on both strands of the DNA for all but 23 residues, where it proved technically difficult. Sequencing each strand of the DNA is a valuable confirmation of the sequence and is probably essential for 100% accuracy.

In addition to the above checks care was taken to overlap the sequence data derived from different restriction fragments in the noncoding region and also in the coding region if the amino acid sequence was uncertain. In the coding region no overlap was sought across the fragments labelled at the *Pst*I site (residue 477) or the *Alu*I site (1,042), because the amino acid sequence was reliable. By contrast, an *Alu*I fragment overlapping the *Hae*III site (818) was sequenced because here the amino acid sequence was not established. In the noncoding

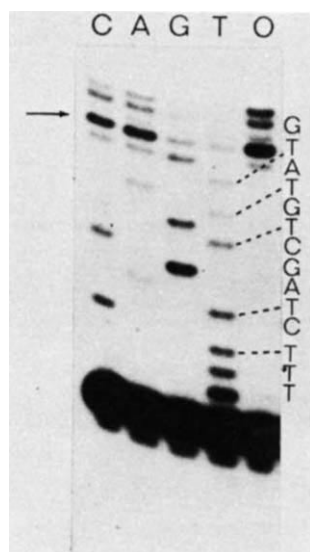


Fig. 3 Sequence of 5' end of mRNA by 'primed synthesis'. C, A, G and T refer to the four separate dideoxynucleoside triphosphate reactions. O is a control containing all four deoxynucleoside triphosphates but with no added dideoxynucleoside triphosphates. The arrow marks the major 'full stop' band while the intense band at the bottom is unused primer. The primer was a ^{32}P -end-labelled 23-long *AluI* fragment (residues 17–39) prepared by ^{32}P -end-labelling an *AluI* digest of 40 μg of pOv230 using a protocol identical to that used for labelling the *PstI* site (see legend to Fig. 1), except that no venom phosphodiesterase was used. After the phosphokinase labelling in 22 μl , 5 μl of the 30% sucrose dye solution was added and the samples loaded directly onto a 1.5 mm $\times$ 25 mm slot of a 40 cm long 12% nondenaturing acrylamide gel in 1 $\times$ Tris-borate-EDTA buffer¹² and electrophoresis carried out for 16 h at 200 V. The 23-long *AluI* fragment ran close to the fast blue dye and was identified by its absence from a control, but otherwise identical digest and labelling, of pMB9; it was eluted in the presence of 20 μg carrier tRNA, but excluding sodium dodecyl sulphate. Before setting up the 'dideoxy' reactions, a tenth of the fragment (2 μl) was boiled with 20 μl water for 2 min to denature it and then mixed with 25 μl of 1.5 mg ml⁻¹ of partially purified ovalbumin mRNA (from S. L. Woo). 4 μl of this primed mRNA was used in a 10 μl reaction volume containing all four deoxynucleoside triphosphates at 0.02 mM and the relevant single dideoxynucleoside triphosphate (from F. Sanger) at 0.1 mM with buffer¹² and reverse transcriptase (from W. Beard). After 45 min at 37 $^{\circ}\text{C}$, 3 μl of a 90% formamide, 0.02 mM EDTA solution containing dyes was added and the sample boiled for 5 min in an open tube to reduce the volume to 3 μl , which was loaded onto a 5 mm wide slot of a 40 cm long 12% thin gel⁴¹ and electrophoresed at 1.3 kV for 1.5 h.

region only the *HinfI* site (1,458) was not overlapped by sequencing an independent restriction fragment. However, in this case the internal 'N' nucleotide of this '5' overhanging' restriction site (G-A-N-T-C) was sequenced on both strands and provided the overlap. This observation made it unlikely that there were two *HinfI* sites very close together.

The sequence near the poly(A) region was less extensively checked because residues 1,730–1,859 had been sequenced previously using 'primed synthesis' on the mRNA directly¹². However, we established that pOv230 included all the known poly(A)-adjacent sequences of the mRNA by sequencing a *HinfI* fragment labelled at residue 1,732. The tentative sequence previously proposed for the region of 131–191 residues from the poly(A) was corrected (nine changes), eliminating two postulated restriction sites. Thus no apparent mutation or rearrangement of sequences occurred during cloning so that the plasmid represents an exact copy of the mRNA. Despite the care taken to ensure accuracy the possibility of occasional errors remains.

Sequence at the 5' end of the mRNA

The sequence of pOv230 extended into the 5' noncoding region, but there was no way of assessing from the sequence itself how close it was to the 5' end of the mRNA. These additional 12 nucleotides were defined by using primed synthesis on the mRNA template with reverse transcriptase. Similar methods have been used to define the 5' noncoding regions of globin mRNA²⁵. A chain terminator method²⁶ (P. Hamlyn *et al.*, in preparation) using dideoxynucleoside triphosphates was used to define this sequence (Fig. 3). Residue 1 (an A) was deduced from a 'partial minus' reaction²⁵ (results not shown). In these various reactions with reverse transcriptase some minor synthesis invariably occurred for 1–3 residues beyond the major 'full-stop'²⁵ position. We have not established whether this represents a length microheterogeneity at the 5' end of the mRNA, or whether it reflects errors in copying due to the presence of the cap structure. The presence of the m⁷Gppp cap structure was assumed from translation studies²⁷. We have not assessed whether N⁶ methylation occurs at residue 1 as in the rabbit globin mRNAs²⁸.

Significance of the mRNA sequence

The 1,859 residue long ovalbumin mRNA is the third eukaryotic mRNA to be completely sequenced and makes an interesting comparison with the completed rabbit²⁹ and human^{25,30} β -globin sequence as well as the partial sequence available for α -globin²⁵, immunoglobulin (P. Hamlyn *et al.*, in preparation) and other mRNAs. Like all eukaryotic mRNAs it has more nucleotides than are needed to code for its specific protein, but the length of the 3' noncoding region is remarkable. Specifically, ovalbumin mRNA has a 64-long 5' noncoding region, a 1,158-long coding region, and a 637-long 3' noncoding region. The 637 noncoding residues at the 3' end are more extensive than the entire rabbit β -globin sequence (589 residues).

(1) *5' Noncoding region.* This region is believed to be required for initiation of protein synthesis, but the molecular details of how this is achieved are unknown. The sequence is of comparable length to other 5' noncoding sequences but is otherwise quite dissimilar in sequence. It seems that only the cap and A-U-G are conserved sequence features for all the 17 known 5' noncoding sequences. This has led to the suggestion that apart from the cap, which is known to augment initiation, A-U-G itself is the only important general signal for initiation in higher organisms⁴⁴.

It is notoriously difficult to predict secondary structure models from primary sequence studies, but nevertheless a convincing 'hairpin loop' can be drawn close to the cap (Fig. 4). This is a highly stable structure³¹ ($-\Delta G$ (25 $^{\circ}\text{C}$) = 11.4 kcal) and additional data for its existence was evident in the sequence determination of this region (Fig. 4 legend). A tight hairpin loop near the cap is not a conserved feature in all mRNAs, but it might have a special, if as yet unknown, significance for ovalbumin mRNA. However, we could speculate that this hairpin loop exists to bring the 'cap' and the A-U-G close together in the tertiary structure of the mRNA, despite the fact that these are 65 residues apart in the primary structure. The remainder of the 5' noncoding region may also participate in local hairpin loops but these are of marginal stability and several alternatives are possible, which cannot yet be distinguished. It may be noted that there are no A-U-G triplets other than the initiation triplet in this region suggesting that the first A-U-G from the cap might specify initiation. But, as this is not true of other known 5' noncoding sequences⁴⁴, the significance of this is doubtful.

(2) *Coding regions.* The entire mRNA sequence, even excluding the poly(A) region, is rich in A+U (A+U=58.1%) and the coding region reflects this (A+U=55.3%). This imposes no severe constraints on the amino acid sequence because of

redundancies in the genetic code. All codons are used at least once (Fig. 5) with the exception of C-C-G (proline), G-C-G (alanine) and C-G-U, C-G-A and C-G-G (arginine). These codons all possess the doublet C-G, which is rare in eukaryotic DNA³² and in globin mRNA²⁹. C-G doublets are not entirely excluded since there are 12 in the coding region. Nevertheless, this is a significantly lower number than is found for any of the other 15 doublets. The reason for the rarity of (5') C-G (3') as opposed to the other 15 possible doublets is not known but may be correlated with *N*⁵ methylation of the cytosine residues in the DNA³³. Other codons which are rarely used are C-U-A (once, leucine) and C-C-C (once, proline). Nonrandom use of codons has been noted before in bacteriophage Φ X DNA³⁴ and in rabbit β -globin mRNA²⁹ but the reason is again unknown.

(3) *3' Noncoding region.* This region is most obviously characterised by its length of 637 nucleotides including the U-A-A terminator, and is by far the longest known 3' noncoding region. It compares, for example, with 95 residues in rabbit β -globin and 211 residues in a mouse κ immunoglobulin light chain (P. Hamlyn *et al.*, in preparation). A detailed examination of this sequence reveals several other features which distinguish it from the coding region. For example, 'runs' of a single base (for example A₅, A₈, G₅) are more frequent than would be expected for a random nucleotide sequence of equivalent overall composition and length. Another feature is the presence of short repeats or near repeats: G₄-U-A₄ (residues 1,285–1,293) is repeated with one base change as G₅-A₄ (1,507–1,515) and G-A₅-U-G-C-A is also exactly repeated (1,471 and 1,511). The A-rich sequence immediately following the U-A-A terminator is particularly striking and is repeated with one change in the A₈ sequence (1,351). There are many other shorter runs and repeats. There are no C-G doublets.

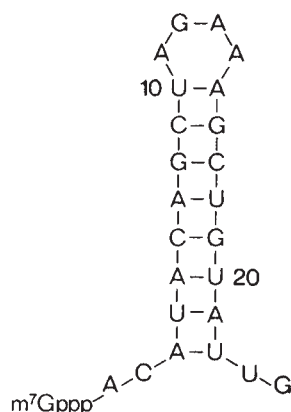


Fig. 4 Hairpin loop near cap. The sequence determinations in this region showed severe 'compression' (close spacing) at the T-G-T sequence when the acrylamide gel of Fig. 3 was run 'cold', that is at 250 V, 16 h on a standard 1.5 mm thick gel, rather than 'hot' as described in the legend to Fig. 3. 'Compression' invariably correlates with hairpin loop structures, and so provides positive evidence for their existence¹².

A comparison with the same region of other mRNAs reveals certain similar features which have not previously been emphasised. For example, rabbit α -globin mRNA³⁵ contains a C₈ sequence while the human α -globin contains three near repeats of the hexanucleotide G-G-G-C-C-C²⁵. An immunoglobulin light chain mRNA also contains an unusual localised repeat based on the triplet U-C-C (P. Hamlyn *et al.*, in preparation). Moreover, longer areas covering 30 nucleotides of partial repeat sequence were noted in the human β -globin mRNA³⁶. Simple repeat sequences are also characteristic of nontranscribed DNA under no selective pressure³⁷, such as parts of the 5S DNA 'spacer' of *Xenopus*³³ and also satellite DNA.

		Second position →					
		U	C	A	G		
First position ↓	U	Phe 11 Leu 3	Ser 10 7	Tyr 3 5	Cys 5 1	U 1 C 1	
	C	Leu 8 1	Pro 6 7	His 4 3	Arg 0 2	U 0 C 0	
	A	Ile 6 3	Thr 4 7	Asn 10 7	Ser 6 9	U 6 C 9	
	G	Val 7 5	Ala 9 16	Asp 8 18	Gly 3 6	U 3 C 6	
		10	0	15	2	G 10 A 2	

Fig. 5 Codon usage in ovalbumin mRNA.

The presence of these homopolymeric and repeat sequences suggests that the 3' noncoding region of ovalbumin mRNA, and indeed other mRNAs, may not be a region with a precise sequence-dependent function. Rather these regions may have similar evolutionary origins through multiple repetitions of simple sequences. This could explain the variable length and sequence features in all of the known 3' noncoding regions—a variability that would be surprising if special functions were to be ascribed to these regions. It is also possible that there is a limited sequence variation in this region of the mRNA which would not be revealed here as the sequence was determined mainly from a single clone. Despite this, there may be evolutionary constraints on the 3' noncoding sequence of a particular mRNA because it has some property, for example the overall shape of the mRNA, which is not sequence-specific. This could explain the observed 80% homology of the noncoding regions in rabbit and human β -globin mRNA³⁶.

One sequence, the hexanucleotide A-A-U-A-A-A (residues 1,841–1,846) is conserved in different mRNAs and occurs about 20 residues away from the poly(A)¹⁴ but its function remains unknown. Given the evolutionary explanation for the 3' noncoding region and its position near the 3' end of the mRNA, the strong possibility emerges that it is related to the termination of transcription or to a subsequent processing of the mRNA. However, A-A-U-A-A-A occurs elsewhere in the coding region of ovalbumin mRNA (223–228), suggesting that it alone may be a necessary but not a sufficient signal.

Protein sequence

The amino acid sequence predicted here is in complete agreement with an almost completed analysis using classical direct sequence analysis of ovalbumin (A.D.N. *et al.*, in preparation). It is clear, therefore, that the sequence shown above represents the major ovalbumin protein corresponding to one of the known genetic variants²⁴. The other allelic form can be explained by a point mutation of residue 998 from A to G, giving an aspartic acid instead of an asparagine residue. Ovalbumin is both glycosylated and phosphorylated at the positions shown (Fig. 2), and the position of the single 'intrachain' disulphide bond is established in part³⁸.

Ovalbumin is a secreted protein. Most secreted proteins have a transient N-terminal hydrophobic region of sequence or 'signal peptide'³⁹ which is removed during the secretion process. But translation studies²⁰ have shown that ovalbumin lacks such a signal sequence. The nucleotide sequence that we have established substantiates this observation, as an A-U-G codon immediately precedes the triplet coding for the first amino acid in ovalbumin. Presumably a somewhat specialised mechanism exists for ovalbumin secretion.

We thank R. Staden for assistance with computer data storage and analysis and the MRC, the Baylor Center for Population Research and Reproductive Biology, and the Keasbey Memorial Foundation for financial support. D.G. was recipient of an EMBO fellowship.

Received 12 April; accepted 11 May 1978.

1. Shrader, W. & O'Malley, B. W. in *Receptors and Hormone Action 2*, (eds O'Malley, B. W. & Birnbaumer, L.) 189-224 (Academic, New York, 1978).
2. O'Malley, B. W., McGuire, W. J., Kohler, P. O. & Korenman, S. G. *Rec. Prog. Horm. Res.* **25**, 105-160 (1969).
3. O'Malley, B. W. & Means, A. R. *Science* **183**, 610-620 (1974).
4. Breathnach, R., Mandel, J. L. & Chambon, P. *Nature* **270**, 314-319 (1977).
5. Doel, M. T., Houghton, M., Cook, E. A. & Carey, N. H. *Nucleic Acids Res.* **4**, 3701-3713 (1977).
6. Lai, E. C., Woo, S. L. C., Dugaiczky, A., Catterall, J. F. & O'Malley, B. W. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
7. Weinstock, R., Sweet, R., Weiss, M., Cedar, H. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1299-1303 (1978).
8. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097-1108 (1977).
9. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725-729 (1978).
10. Brack, C. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5652-5656 (1977).
11. Gilbert, W. *Nature* **271**, 501 (1978).
12. Brownlee, G. G. & Cartwright, E. M. *J. molec. Biol.* **114**, 93-117 (1977).
13. Cheng, C. C. *et al. J. molec. Biol.* **107**, 527-547 (1976).
14. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
15. Monahan, J. J., Woo, S. L. C., Liarakos, C. D. & O'Malley, B. W. *J. biol. Chem.* **252**, 4722-4728 (1977).
16. McReynolds, L. A., Catterall, J. F. & O'Malley, B. W. *Gene* **2**, 217-231 (1977).

17. Humphries, P., Cochet, M., Krust, A., Gerlinger, P., Kourilsky, P. & Chambon, P. *Nucleic Acids Res.* **4**, 2389-2406 (1977).
18. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
19. Staden, R. *Nucleic Acids Res.* **4**, 4037-4051 (1977).
20. Palmer, R. O., Gagnon, J. & Walsh, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 94-98 (1978).
21. Thompson, E. O. P., Sleight, R. W. & Smith, M. B. *Aust. J. biol. Sci.* **24**, 525-534 (1971).
22. Lee, Y. C. & Montgomery, R. *Archs Biochem. Biophys.* **97**, 9-17 (1962).
23. Milstein, C. P. *Biochem. J.* **110**, 127-134 (1968).
24. Wiseman, R. L., Fothergill, J. E. & Fothergill, L. A. *Biochem. J.* **127**, 775-780 (1972).
25. Baralle, F. E. *Nature* **267**, 279-281; *Cell* **12**, 1085-1095 (1977).
26. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
27. Sharma, O. K., Hruby, D. E. & Beezley, D. N. *Biochem. biophys. Res. Commun.* **72**, 1392-1398 (1976).
28. Lockard, R. E. & RajBhandary, U. L. *Cell* **9**, 747-760 (1976).
29. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. *Cell* **10**, 571-585 (1977).
30. Wilson, J. T. *et al. Nucleic Acids Res.* **5**, 563-581 (1978).
31. Tinoco, I., Jr *et al. Nature new Biol.* **246**, 40-41 (1973).
32. Subak-Sharpe, J. H. *Br. med. Bull.* **23**, 161-168 (1967).
33. Miller, J. R., Cartwright, E. M., Brownlee, G. G., Fedoroff, N. & Brown, D. D. *Cell* **13**, 717-725 (1978).
34. Sanger, F. *et al. Nature* **265**, 687-695 (1977).
35. Proudfoot, N. J., Gillam, S., Smith, M. & Longley, J. I. *Cell* **11**, 807-818 (1977).
36. Proudfoot, N. J. *Cell* **10**, 559-570 (1977).
37. Smith, G. P. *Science* **191**, 528-535 (1976).
38. Fothergill, L. A. & Fothergill, J. E. *Biochem. J.* **116**, 555-561 (1970).
39. Devillers-Thiery, A., Kindt, T., Scheele, G. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5016-5020 (1975).
40. Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82**, 314-316 (1977).
41. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107-110 (1978).
42. Bolivar, F., Rodriguez, R. L., Betlach, M. C. & Boyer, H. W. *Gene* **2**, 75-93 (1977).
43. Roberts, R. J. *Crit. rev. Biochem.* **4**, 123-164 (1976).
44. Baralle, F. E. & Brownlee, G. G. *Nature* (in the press).

letters to nature

A Near-Eastern sighting of the supernova explosion of 1054

ASTRONOMERS have long been puzzled by the apparent failure of observers in Europe and the Near East to see and report the Crab supernova of 1054. However, we now present evidence of such a sighting.

On 4 July 1054 a supernova explosion was observed and recorded in several reports by astronomers in China and Japan. For example, the report in the *Sung-shih* (Astronomical Treatise, Chapter 56) says, "First year of the Chih-ho reign period, fifth month, (day) *chi-ch'ou* (a 'guest star', that is, a nova) appeared approximately several inches (*k'o-shu-ts'un*) to the south-east of *T'ien-kuan* (ζ Tau). After a year and more it gradually vanished."^{1,2} This remarkable observation was first brought to light in the west by Biot³ in 1846. Lundmark⁴ in his list of suspected new stars recorded in old chronicles, published in 1921, notes the proximity of this object to NGC1052—the Crab nebula. Although Duncan⁵ measured the expansion rate of the filaments of the Crab in the same year (1921) and found that a simple linear extrapolation of their motion with constant velocity implied an origin about 900 years earlier, it was not until 1928 that Hubble⁶ first drew the connection between the expansion rate and the Chinese observation to identify the Crab nebula with the 'guest star' observed by the Chinese in 1054. The subsequent historical researches^{1,7,8} have mainly served to strengthen this identification (but see ref. 9 for some reservations).

The Chinese records indicate that the 'guest star' was brighter than Venus, giving it a probable visual magnitude of -4 or -5 at maximum. It was visible in the day time for 23 days, and continued to be visible in the night sky for perhaps 6 months. It is therefore puzzling that no reference to this event has been found in European or Arabic records of the

time^{6,10-12}. (A possible record in certain south-west Indian petroglyphs has been proposed by Miller¹³, and further supported by Brandt¹⁴. However, see Ellis¹⁵ and S. H. Koenig (in preparation) for some cautionary remarks.) This led to the conclusion, as expressed, for example, by the historian of science George Sarton, that "The failure of medieval Europeans and Arabs to recognise such phenomena was not due to any difficulty in seeing them, but to prejudice and spiritual inertia connected with the groundless belief in celestial perfection."¹⁶ However, since they quite accurately reported the position and other properties of another supernova which occurred in 1006, only 48 years earlier¹⁷, as well as many other astronomical phenomena, such as Halley's comet (witness the Bayeux tapestry), and meteorological phenomena, such as aurorae, this has never seemed to be a very convincing explanation. Others have suggested that the whole of the Middle East and Europe was clouded out for 6 months, an even more dubious proposition. We present here for the first time what we believe to be an eyewitness account of a sighting of the 1054 supernova explosion in the Near East.

The event and its supposed sequelae were mentioned by Ibn Buṭlān, a Christian physician of Bagdad, who lived in Cairo for 3 years until late in 1052 or early in 1053. Then, probably as the result of a famine, he left for Constantinople where he stayed for about a year. His original report was reproduced by Ibn Abī Ūṣaybi'a (hereafter referred to as I.A.U.) in the biography of Ibn Buṭlān, which he includes in his *Uyūn al-Anbā'*, a biographical encyclopaedia of physicians, originally composed around 1242. Ibn Buṭlān was not a professional astronomer or astrologer but, in the tradition of Hippocrates and Galen, he, like other physicians of his time, was concerned with the presumed connection between cosmic and telluric happenings and diseases and other natural catastrophes affecting the life and health of man. The report¹⁸ reads as follows: