Nucleotide Sequence Analysis of RNA Synthesized from Rabbit Globin Complementary DNA

(reverse transcriptase/RNA polymerase/globin mRNA)

RAYMOND POON*, GARY V. PADDOCK*, HOWARD HEINDELL*, PHILIP WHITCOME*, WINSTON SALSER*, DAN KACIAN†, ARTHUR BANK‡, ROBERTO GAMBINO‡, AND FRANCESCO RAMIREZ‡

* Department of Biology and Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90024; and
† Institute of Cancer Research, and ‡ Departments of Medicine and Human Genetics and Development, Columbia University College of Physicians and Surgeons, New York, N.Y. 10032

Communicated by Jacob Furth, June 13, 1974

ABSTRACT Rabbit globin complementary DNA made with RNA-dependent DNA polymerase (reverse transcriptase) was used as template for *in vitro* synthesis of ³²P-labeled RNA. The sequences of the nucleotides in most of the fragments resulting from combined ribonuclease T₁ and alkaline phosphatase digestion have been determined. Several fragments were long enough to fit uniquely with the α or β globin amino-acid sequences. These data demonstrate that the cDNA was copied from globin mRNA and contained no detectable contaminants.

The RNA-dependent DNA polymerase (1, 2) from avian myeloblastosis virus (AMV) has been shown to synthesize a DNA copy from a variety of RNA templates (3-12). The single-stranded complementary DNA (cDNA) synthesized from globin mRNA in the presence of actinomycin D is well characterized (7-10). The DNA product has a size similar to that of the RNA template as determined by alkaline sucrose gradient centrifugation (13), and it reanneals specifically to the RNA template. This cDNA has been used to test for the presence of globin sequences in heterogeneous nuclear RNA (14), to measure the relative amounts of α and β globin mRNA in reticulocytes from thalassemics (15, 16), to study control of transcription of various tissue-specific chromatin *in vitro* (17, 18), and to determine the number of globin genes (19, 20).

Transcripts of globin mRNA should also be of great value in studying the coding properties of eukaryotic messages. *In vitro* synthesis offers the considerable advantages of a very high specific activity and nearest neighbor label transfer, which greatly aid in determining the nucleotide sequences. Detailed knowledge of the mRNA sequence should answer fundamental questions about the nature of the signals for initiation and termination of protein synthesis, about the signals for addition of poly(A) sequences, about the tertiary structure of the mRNA molecule, and about large untranslated sequences that are present.

We have recently begun sequencing RNA transcripts of the globin cDNA made *in vitro* with *Escherichia coli* RNA polymerase. Some preliminary ribonuclease T₁ fragment sequences deduced from pancreatic ribonuclease digestions and nearest neighbor transfer of label were published previously (21). In the present communication we report the results of much more detailed analyses of the ribonuclease T₁ fragments by digestion with ribonuclease U₂ and by partial digestions with spleen phosphodiesterase as well as digestion with pancreatic ribonuclease. We have continued to find that the fingerprints obtained are highly reproducible with good agreement between

the results of different syntheses from the same batch of cDNA or from cDNA preparations made at different times. Of the sequences published earlier, most have been confirmed, and some have been revised as a result of this more detailed analysis. Several long new sequences have been established and found to fit with unique sites in the known rabbit globin amino-acid sequences. From the positions of these unique fits, it is clear that large portions of the globin mRNAs may be sequenced using this approach.

METHODS

Isolation of rabbit globin mRNA, purification of RNA-dependent DNA polymerase from avian myeloblastosis virus, and synthesis of cDNA from globin mRNA have been described (22, 23, 9). *E. coli* RNA polymerase, isolated by the procedure of Burgess (24), was the kind gift of Cindy Clemenson and Dr. David Eisenberg.

In Vitro RNA Synthesis. Conditions were similar to those of Burgess (24). Reactions (100 μ l of final volume) contained 4 μ moles of Tris·HCl (pH 7.9), 1 μ mole of MgCl₂, 0.01 μ mole ethylenediaminetetraacetate (EDTA), 0.01 μ moles of dithiothreitol, 0.04 μ moles of potassium phosphate (pH 7.5), 10 μ g of RNA polymerase, and 1 μ g of globin cDNA, and 16 nmoles of each ribonucleoside triphosphate, one labeled with ³²P in the alpha position (specific activity about 100 Ci/mmole). The reaction mixture containing GTP was preincubated at 37° for 5 min before the other nucleotides were added. Samples were taken for determination of acid-precipitable radioactivity. Synthesis was usually stopped after 45 min. The material was heat-denatured and desalted on a Bio-Gel P-60 column containing 0.01 M triethylammonium carbonate (pH 9.5).

Sequence Determination. Procedures for enzymatic digestions and one- and two-dimensional electrophoretic analysis of oligonucleotides have been extensively described by Sanger and colleagues (25-27).

RESULTS

***In vitro* RNA synthesis**

The single-stranded cDNA template with its stretch of deoxythymidylic acid residues at the 5'-end poses unique problems for *in vitro* synthesis. Incorporation of AMP is 5- to 10-fold greater than that of other nucleotides and large ribonuclease T₁ and pancreatic ribonuclease-resistant fragments are found in digests of material synthesized with [α -³²P]ATP. This synthesis is not inhibited by inorganic phosphate (28), does

Abbreviation: cDNA, DNA complementary to RNA.

FIG. 1. Fingerprints of combined ribonuclease T_1 and bacterial alkaline phosphatase digest of RNA synthesized *in vitro* from globin cDNA. The labeled precursors are (A) [α - 32 P]GTP, (B) [α - 32 P]UTP, and (C) [α - 32 P]CTP; (D) is a line drawing of (B).

not occur with bacteriophage M13 DNA or without template, and can be reduced 2-fold by short preincubation with GTP before addition of the other nucleoside triphosphates. These results suggest that a substantial amount of poly(A) is synthesized by polymerase molecules initiating directly off the oligo(dT) sequences.

Unlabeled poly(A) does not interfere with the fingerprint analysis. When [α - 32 P]ATP is the radioactive precursor, it is necessary to remove poly(A) sequences because commercial bacterial alkaline phosphatase and bovine pancreatic ribonuclease contain activities that cleave them. Such cleavages yield polydisperse [32 P]oligoadenylate which can mask the graticule pattern. By treating alkaline phosphatase with diethylpyrocarbonate (29) and by using pancreatic ribonuclease under high salt conditions (30, 31), one can eliminate degradation of poly(A) tracts. Alternatively, poly(A) sequences may be removed by passing the enzymatic digests through an oligo(dT)-cellulose column (32).

The KCl concentration in the transcriptase reaction influences the product length, smaller transcripts being produced at higher salt concentrations with single-stranded DNA templates (33). In 0.15 M KCl, globin RNA synthesized *in vitro* has an average size of 70 nucleotides, whereas when KCl is absent, the product size is comparable to that of the cDNA template. Similar fingerprints are obtained from material synthesized in either high or low salt, as has been found by others using simian virus 40 DNA template (34). This suggests that the same regions of the cDNA are transcribed in both cases.

Two-dimensional electrophoresis

Initial indications of the fidelity of the reverse transcriptase reaction derive from the complexity and reproducibility of the fingerprint pattern with different batches of cDNA. Fig. 1 shows fingerprints of combined ribonuclease T_1 and alkaline phosphatase digests of *in vitro* synthesized globin RNA in which the radioactive precursors are [α - 32 P]GTP (Fig. 1A), [α - 32 P]UTP (Fig. 1B), [α - 32 P]CTP (Fig. 1C), respectively. Except for expected differences due to the labeling (for example, the graticule containing no U residues would not be

visible when [α - 32 P]UTP is the radioactive precursor), the fingerprints are remarkably similar. Subsequent analyses of the fragments confirm their reproducibility. The "complexity" of the fingerprint can give a rough estimation of the size of the sequence represented. As discussed by Salser *et al.* (35), the number of spots in the upper half of the fingerprint in Fig. 1, when compared to that of a well-characterized fingerprint of similar complexity such as that of *E. coli* 16S ribosomal RNA (36), can be used to estimate the complexity of the *in vitro* globin RNA relative to 16S rRNA. These calculations (21) suggest that most of the sequences of α and β globin mRNA are found in the RNA synthesized from the cDNA template.

Sequences of the fragments

Tables 1 and 2 list the sequences of residues in the oligonucleotides produced by RNase T_1 digestion that have been determined to date. They were deduced from analyses of complete enzymatic digestions with ribonuclease U_2 or pancreatic ribonuclease, and from partial digestions with spleen phos-

TABLE 1. Sequences of T_1 oligonucleotides composed of less than 7 residues

Fingerprint spot number	Sequence	Fingerprint spot number	Sequence
2	GCUG	19	GCUUG
3	GAUG	20	GUAUG
4	GCCUG	21a	GCUCUG
5	GACUG	21b	GUCCUG
	GCAUG	23	GAUUG
	GAUCG	24	GACUUG
6	GAAUG		GAUCUG
7	GCCCUG	25	GUUAAG
8	GCACUG	37	GUUUG
9	GACCUG	38	GUUAUG
10	GAACUG	39	GUCUUG
16	GUUG	49	GUUUUG
17	GUCUG		

TABLE 2. Sequences of T_1 oligonucleotides composed of 7 or more residues

Fingerprint spot number	Sequence	Amino-acid fits*		Numbers of fits
		Alpha chain	Beta chain	
11	GCCAAUG	None	ANA (51, 138)	2
12	GACCCUG	VTL (107), STL (81), DPV (94)	DPE (99)	4
14	GCCACCUG		SHL (76)	2
15	GCAAAUG		ANA (51, 138)	2
22	GCUCCUG	KLL (99)	RLL (30, 104)	3
27a	GUCACUG	SHC (102)	VTA (11)	3
		ASL (123)		
27b	GCUCAUG	AHG (57)	AHG (62)	2
27b'	GCAUCUG	GHL (71)	VHL (1)	3
			SHL (76)	
28	GAUCCUG	DPB (94)	DPE (99)	2
31	GUCCCACUG	LSH (101)	None	1
33	GCCUAUCAG	None	AYQ (129)	1
40	GUUCCUG	MLF (32)	None	2
		KFL (127)		
42	GUCCUCUG	None	LSSA (48)	2
			LSSE (3)	
44	GACUUCUG	None	RLL (30, 104)	2
45	GCACCUUUG	None	GTFA (83)	1
47	GAAUUCACCCUG	EFTPA (116)	None	1
48	GAAUUUCAAG	VNFK (96)		1
50	GUUAUUG	KLL (99)	RLL (30, 104)	4
			VIV (111)	
51	GUCUUUG	ASL (123)	None	1
52	GUUCUUCG	None	RFFE (40)	1
53	GUCCUUUG	None	ESFG (43)	1
54	GUCUACUCUCAG	LSTLS (80)	None	1
58c	GUCUCAUUAUUUG	None	LSHHFG (114)	1

* The amino-acid sequences of rabbit globins are based on ref. 40. The single letter code for amino acids used in ref. 40 is used here. The numbers in parentheses are the positions of the first indicated amino acid in the globin chains.

phodiesterase. Radioactive label was incorporated in only one nucleoside triphosphate at a time, giving additional sequence information through nearest neighbor labeling.

Oligonucleotides of seven or more bases were checked by computer for correspondence with the amino-acid sequences of the α and β globin chains (Table 2). Most were found to fit within the three possible reading frames into two or more positions in either or both chains; however, some sequences are sufficiently long so that they are unlikely to have any "chance" fits with the amino-acid sequence. Among sequences with single fits are those for spots 31, 47, 48, 51, and 54, which fit the α chain, and spots 33, 45, 52, 53, and 58c, which fit the β chain.

To assess the significance of the observed fits with the amino-acid sequences, we have checked the fits of random nucleotide sequences of various sizes. Fig. 2 shows the correlation between oligonucleotide length and number of fits with the globin amino-acid sequences. As expected, short oligonucleotides fit with high frequencies whereas oligonucleotides of increasing size show rapidly decreasing probabilities of a chance fit. The average number of fits is > 1.0 for lengths up to 7 nucleotides and less than 0.1 for lengths greater than 9 nucleotides. The average number of fits obtained with the real fragments of each size class are also plotted in Fig. 2. Notice that the real heptanucleotides fit the amino-acid sequence at an average of 2.54 places whereas the random ones fit only 1.45 times. This is exactly the expected result, since a real sequence should in general have the same number of chance fits as a random sequence *plus* its real fit. For the octa- and nona-

nucleotides, we see again that real sequences transcribed from globin cDNA have, on the average, one more fit than random nucleotide sequences of the same length.

Comparisons with known amino-acid sequences can also be used to resolve ambiguities in the nucleotide sequence data (37). We have had to rely on this aid in only one case, that of spot 58c for which the sequence G(UCUCA, UCA)UUUUG was determined. The comma indicates uncertainty in the order of the two ribonuclease U_2 products, UCUCA and UCA. Of the two possibilities, only the sequence GUCUCAU-CAUUUUG would fit the known amino-acid sequences of either the α or β chains. Because the fragment sequence is long, it is statistically very improbable that this is a chance fit. We are currently trying to independently confirm this sequence assignment by conventional methods.

Two sequence pairs with unique fits are sufficiently close together that we can deduce longer nucleotide sequences from the amino-acid data (Table 3). In one case we can construct a 19 nucleotide sequence without ambiguities, and in the other a 32 nucleotide sequence with uncertainties in three positions. We have also identified fragments that could fit the gaps shown in the 32 nucleotide sequence, but they are too small for meaningful fits. We will have to sequence overlapping fragments to provide the necessary information.

DISCUSSION

Due to the large intracellular phosphate pools, it is difficult to prepare ^{32}P -labeled eukaryotic mRNAs *in vivo* with the very high specific activities needed for rapid sequencing. Moreover,

such preparations are often contaminated with labeled rRNA fragments. We pointed out earlier (21) that the use of reverse transcriptase should circumvent both of these problems for poly(A)-containing mRNAs. We routinely obtained 10^8 cpm of ^{32}P -labeled product from $1\text{ }\mu\text{g}$ of cDNA template, and the use of oligo(dT) primer for cDNA synthesis insures that only poly(A)-terminated mRNAs will be copied. For sequencing studies of this sort to give meaningful results, it is essential that the reverse transcriptase work with high fidelity. Our earlier data (21) ruled out the possibility that the enzyme makes significant numbers of random mistakes, since this would have resulted in fingerprint patterns having a high background of many faint spots. Our fingerprints have routinely been very "clean" except for occasional faint spots resulting from incomplete phosphatase treatment. Not only does this rule out significant levels of random errors, but, equally important, it suggests that the globin cDNA is not contaminated with cDNA copies from other messengers or rRNA. There remained the possibility that reverse transcriptase might systematically make specific errors, always inserting a particular wrong base at certain positions in the sequence; however, the data presented here indicate that such mistakes are not made. If such errors had occurred in the longer oligonucleotides, they should frequently have been detected as sequences that almost, but not quite, fit the amino acid sequence. Only one possible case of this kind has appeared. Spot 48 (GAAUUCAAG) fits the amino-acid sequence valine-asparagine-phenylalanine-lysine beginning at position 96 of the α -chain. Experiments using isoacceptor lysyl-tRNAs have led to some controversy whether the lysine (position 99) is coded for by AAA or AAG (38, 39).

The only apparent drawback to the use of reverse transcriptase in this way to sequence eukaryotic mRNAs may be the possibility of incomplete copying either during the RNA to cDNA step or the cDNA to RNA step or both. If some information is lost, it may be possible to alter reaction conditions so that all regions are ultimately represented. Even if this cannot be achieved, the technique still permits rapid sequencing of considerable portions of the RNA.

Note Added in Proof. Marotta *et al.* (41) have recently reported complete sequences for nine fragments from human globin mRNA

FIG. 2. Graph of the number of fits with the amino-acid sequences of randomly generated nucleotide sequences as a function of their chain length. In order that these would correctly simulate the RNase T_1 fragments actually obtained, we chose those sequences beginning and ending with G residues. For the internal sequences, we used a random sequence generating function which would produce runs of A, U, and C residues with the same average base composition as the hemoglobin RNase T_1 fragments of the same chain lengths. The sample sizes for random sequences are 50 for each chain length. ●—● represents randomly generated sequences; ○—○ represents actual nucleotide sequences.

* Represents the average number of fits for the four real fragment sequences ranging from 10 to 14 nucleotides in length.

and partial sequences for eight more. In four cases fragments they determined have amino-acid fits corresponding to those we have observed in rabbit globin mRNA.

This permits the first direct inspection of evolutionary changes in the nucleotide sequences of mammalian mRNAs. In the case of β chain positions 83-86 all nine bases are the same for rabbits and humans. For β 114-119, three bases out of 13 are different. Regarding the α chain positions 123-125, we find two fragments with unique fits: spot 51, sequence (G)UCUUUG, and spot 26 which we have recently shown to be (G)CCUCCUG. Assuming that the larger of these has the correct fit, we conclude that rabbit and human mRNAs agree at seven out of nine base residues. We have also recently determined that our spot 34c is (G)-AACUUCAG, which has a unique fit with rabbit β 101-104. Here nine out of nine bases agree.

We can now directly compare the rates of fixation of mutations that have resulted in amino-acid substitutions and of mutations that have no such effect. Out of 40 nucleotides "tested" by the

TABLE 3. Composite nucleotide sequences derived from amino-acid sequence information

T ₁ fingerprint numbers of sequences contained in composite sequence	Hemoglobin mRNA source	Derivation of composite sequences												Number of nucleotides in composite sequence					
47 and 51	Alpha (a)	(b)	Ser	116	Phe	Thr	Pro	Ala	Val	His	123	Ser	Leu	Asp					
			U C X	Glu							Ala					U C X	C U X		
			A G U	G A A							G C X					A G U	U U A	G A U	
			C	G							C					C	G	C	
(c)	G A A	U U C	A C C	C C U	G C X	G U X	C A U	G C G	U C U	U U G	G A	32							
52 and 53	Beta (a)	(b)	Gln	40	Phe	Phe	Glu	Ser	Phe	Gly	Asp								
			C G X	Arg											U C X				
			A G A	A G A											A G U	G A A	A G U	G G X	G A U
			G	G											C	C	G	C	C
(c)	G G	U U C	U U C	G A G	U C C	U U U	G G					19							

Lines (a) are the known amino-acid sequences (40); lines (b) are the corresponding possible mRNA sequences, derived from the genetic code; lines (c) are the composite sequences, with the T₁ oligonucleotide sequences underlined.

above data there were four base changes that resulted in no amino-acid substitution. From this and the total number of amino-acid differences between rabbit and human globins, we may compute that base substitutions that cause no amino-acid change are about five times as likely to escape elimination by natural selection. This involves the simplifying assumptions that about one third of all possible base changes cause no amino-acid substitutions and that the 40 nucleotides tested in this preliminary calculation are representative. The rate of fixation of "neutral" mutations thus calculated is similar to that which may be estimated from comparison of fibrinopeptide A amino acid sequences of rabbits and humans.

We hope that there will soon be a much larger body of nucleotide sequence data common to rabbit and human globin mRNAs that will warrant more exact calculations. More important, such comparisons of the globin mRNA sequences may help us to find critical sequences where the lack of base changes will indicate the presence of important secondary structures or essential "signal" sequences.

We would like to thank Ron Saltzman and Randy Bronte for their participation in this project as part of an undergraduate laboratory course at UCLA, and also Mrs. Mary Ann Childs for her expert technical assistance. We are indebted to C. Clemenson and D. Eisenberg for their generous gift of *E. coli* RNA polymerase. We would also like to thank Russell Swatek for setting up the computer program for checking nucleotide sequences for their amino-acid fits and the UCLA Campus Computing Network for their assistance. This research is supported by National Institutes of Health Grants CA-15940, GM-18586, GM-14552, GM-19153, and CA-02332, the Virus Cancer Program Contract NO1-CP-33258 of the National Cancer Institute, National Science Foundation Grants GB-40312X and 27388, and the Cooley's Anemia Foundation. W.S. is the holder of a Career Development Award from the National Institutes of Health (GM-70045), A.B. is a Faculty Research Scholar of the American Cancer Society, and G.V.P. is a Celeste Durand Rogers Cancer Research Fellow.

- Baltimore, D. (1970) *Nature* **226**, 1209-1211.
- Temin, H. M. & Mizutani, S. (1970) *Nature* **226**, 1211-1213.
- Spiegelman, S., Watson, K. F. & Kacian, D. L. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2843-2845.
- Aviv, H., Packman, S., Swan, D., Ross, J. & Leder, P. (1973) *Nature New Biol.* **241**, 174-176.
- Zassenhaus, P. & Kates, J. (1972) *Nature New Biol.* **238**, 139-141.
- Diggelman, H., Faust, C. H., Jr. & Mach, B. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 693-696.
- Ross, J., Aviv, H., Scolnick, E. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 264-268.
- Verma, I. M., Temple, G. F., Fan, H. & Baltimore, D. (1972) *Nature New Biol.* **235**, 163-167.
- Kacian, D. L., Spiegelman, S., Bank, A., Terada, M., Metalfora, S., Dow, L. & Marks, P. A. (1972) *Nature New Biol.* **235**, 167-169.
- Baltimore, D., Verma, I. M., Fan, H. & Temple, G. F. (1974) *Viral Replication and Cancer, Proceedings of 2nd Duran Reynolds International Symposium, Barcelona*, eds. Melnick, J. L., Ochoa, S. & Oro, J. (S. A. Calabria, Barcelona), pp. 155-169.
- Taylor, J. M., Faras, A. J., Varmus, H. E., Goodman, H. M., Levinson, W. E. & Bishop, J. M. (1973) *Biochemistry* **12**, 460-467.
- Taylor, J. M., Faras, A. J., Varmus, H. E., Levinson, W. E. & Bishop, J. M. (1972) *Biochemistry* **11**, 2343-2351.
- Gulati, S. C., Kacian, D. L. & Spiegelman, S. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1035-1039.
- Imaizumi, T., Diggelman, H. & Scherer, K. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1122-1126.
- Housman, D., Forget, B. G., Skoultschi, A. & Benz, E. J., Jr. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1809-1813.
- Kacian, D. L., Gambino, R., Dow, L., Grossbard, E., Natta, C., Ramirez, F., Spiegelman, S., Marks, P. A. & Bank, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1886-1890.
- Axel, R., Cedar, H. & Felsenfeld, G. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2029-2032.
- Gilmour, R. S. & Paul, J. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3440-3442.
- Bishop, J. O. & Rosbash, M. (1973) *Nature New Biol.* **241**, 204-207.
- Hanison, P., Hell, A., Birnie, G. & Paul, J. (1972) *Nature* **239**, 219-221.
- Salser, W., Poon, R., Whitcome, P. & Fry, K. (1973) in *Proc. 1973 UCLA-ICN Symposium on Virus Research*, eds. Fox, C. F. & Robinson, W. S. (Academic Press, New York), pp. 545-572.
- Labrie, F. (1969) *Nature* **221**, 1217-1222.
- Kacian, D. L., Watson, K. F., Burny, A. & Spiegelman, S. (1971) *Biochim. Biophys. Acta* **246**, 365-383.
- Burgess, R. P. (1969) *J. Biol. Chem.* **244**, 6168-6176.
- Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373-398.
- Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper and Row, New York), Vol. 12, pp. 751-779.
- Brownlee, G. G. (1972) in *Laboratory Techniques in Biochemistry and Molecular Biology*, eds. Work, T. S. & Work, E. (North-Holland, Amsterdam), Vol. 3, Part 1, pp. 3-265.
- Ohasa, S. & Tsugita, A. (1972) *Nature New Biol.* **240**, 35-41.
- Wimmer, E. (1972) *J. Mol. Biol.* **68**, 537-540.
- Beers, R. F. (1960) *J. Biol. Chem.* **235**, 2393-2398.
- Gabbay, E. J. & Shimshak, R. R. (1968) *Biopolymers* **6**, 255-268.
- Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1408-1412.
- Maitra, U., Lockwood, A., Dubnoff, J. & Guha, A. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 143-156.
- Zain, B. S., Dhar, R., Weissman, S. M., Lebowitz, P. & Lewis, A. M. (1973) *J. Virol.* **11**, 682-693.
- Salser, W., Fry, K., Wesley, R. & Simpson, L. (1973) *Biochim. Biophys. Acta* **319**, 277-280.
- Fellner, P., Ehresman, C. & Ebel, J. P. (1972) *Biochimie* **54**, 853-900.
- Brownlee, G. G., Cartwright, E. M., Cowan, N. J., Jarvis, J. M. & Milstein, C. (1973) *Nature New Biol.* **244**, 236-240.
- Woodward, W. R. & Herbert, E. (1972) *Science* **177**, 1197-1199.
- Rudloff, E. & Hilse, K. (1971) *Eur. J. Biochem.* **24**, 313-320.
- Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md.), Vol. 4.
- Marotta, C. A., Forget, B. G., Weissman, S. M., Verma, I. M., McCaffrey, R. P. & Baltimore, D. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 2300-2304.