Developmental Biochemistry of Cottonseed Embryogenesis and Germination: Changing Messenger Ribonucleic Acid Populations As Shown by Reciprocal Heterologous Complementary Deoxyribonucleic Acid–Messenger Ribonucleic Acid Hybridization†

Glenn A. Galau and Leon Dure III*

ABSTRACT: The concentrations of messenger ribonucleic acids (mRNAs) in three stages in cotton cotyledon ontogeny were examined by total poly(A)+ mRNA hybridization with unfractionated complementary DNA. This study and others show that these RNAs are representative of the total mRNA population. A novel analysis of all reciprocal hybridization reactions detected at least 17 groups of mRNAs, the sequences of which change together in concentration during this developmental period. We have defined an mRNA subset as a group or groups of mRNAs that change together in concentration in a similar fashion, regardless of the abundance of its members. About 11 such subsets are detected, several of which contain at least two groups which change in concentration in parallel. These experiments have identified the same mRNA subsets detected in a companion study of the abundant proteins synthesized in vivo and in vitro during this developmental period and indicate that similar changes occur for less abundant mRNAs as well.

In the preceding paper (Dure et al., 1981), changes in the population of abundant mRNAs that occur in the cotyledons of the cottonseed embryo during embryogenesis and early germination were monitored by in vivo and in vitro protein synthesis. In this paper we have monitored these changes by nucleic acid hybridization, which, in addition to providing information on this subject by totally different means, extends the study to include changes in the less abundant messenger ribonucleic acid (mRNA)[1] classes. The approach we have used in examining changes in sequence concentration of mRNAs through development relies entirely on homologous and heterologous reactions between mRNAs and unfractionated cDNAs. With hybridization of mRNAs and cDNAs from three developmental stages in all possible pairwise combinations, the analysis allowed a detailed description of the changes in sequence concentration. We find good agreement with the changes in the abundant mRNAs deduced by the study of patterns of protein synthesis in vivo and in vitro (Dure et al., 1981) and find that corresponding changes occur for the less abundant mRNAs as well.

Experimental Procedures

The methods used here are described in detail in Galau et al. (1981). Total RNA was prepared from cotyledons of 50-mg (young) embryos, 110-mg (older) embryos, dry seeds (mature embryos), and seedlings germinated 12 and 24 h by standard methods of deproteinization and precipitation of the RNA with 2 M LiCl and 3 M sodium acetate. Poly(A)+ mRNA was prepared by passage of total RNA through poly(U)–agarose, and complementary DNA was synthesized with reverse transcriptase by using as templates poly(A)+ mRNA from young and mature embryos and seedlings germinated 24 h. The cDNA was removed from anti-cDNA by preparative hybridization with poly(A)+ mRNA and recovery of cDNA-mRNA hybrids on hydroxylapatite (Galau et al., 1981).

Considerable effort was made to reduce experimental variation in the hybridization reactions between the three cDNAs and the several poly(A)+ mRNAs. All three cDNAs were synthesized and purified in parallel. In the first experiment hybridizations of all cDNAs with each of the three poly(A)+ mRNAs (from young and mature embryos and seedlings germinated 24 h) were performed simultaneously by using similar RNA and DNA concentrations in 0.38 M NaCl, 50 mM Pipes–NaOH (pH 6.9), 1 mM EDTA, and 0.1% sodium dodecyl sulfate at 70 °C. Several groups of reaction mixtures (each containing reactions incubated over a wide range of C_0t) were assayed for hybridization by treatment with S-1 nuclease followed by binding of the digestion products to DEAE filters (Salzberg et al., 1977; Maxwell et al., 1978), with parallel controls for S-1 digestion and hybrid recovery. All reaction mixtures at a specific C_0t were assayed together. Less than 2% of the S-1-resistant cDNA in all reactions at high mRNA C_0t was attributable to DNA–DNA reassociation, as assayed by the single-column hydroxylapatite procedure (Galau et al., 1981). In the second experiment cDNA synthesized from young embryo RNA and mRNA from older embryos and seedlings germinated 12 h were incubated and assayed together, but independently of those above. Two observations suggest that the reactions conducted in these two experiments with single preparations of RNA and cDNA are typical. First, several independent hybridizations and assays of the same homologous preparations of cDNA and cotyledon mRNA from young embryos or seedlings germinated 24 h gave results identical with those achieved above. Secondly, several preparations of mRNA and cDNA from 24-h germinated embryos also gave identical hybridization kinetics when incubated and

† From the Department of Biochemistry, University of Georgia, Athens, Georgia 30602. *Received November 19, 1980.* This work was supported by grants from the National Institutes of Health, the National Science Foundation, and Cotton, Inc., to L.D. G.A.G. is a recipient of a National Institutes of Health postdoctoral fellowship. This is paper 15 in this series. Paper 14 is Dure et al. (1981).

[1] Abbreviations used: poly(A)+ or A+, polyadenylated; poly(A)–, nonpolyadenylated; cDNA, complementary DNA; poly(U), polyuridylate; C_0t, product of nucleic acid concentration, in moles of nucleotide per liter, and time, in seconds; EDTA, ethylenediaminetetraacetic acid; Pipes, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid); DEAE, diethylaminoethyl; *K*, pseudo-first-order rate constant in units of $M^{-1}\cdot s^{-1}$; NT, nucleotides; mRNA, messenger ribonucleic acid.

0006-2960/81/0420-4169$01.25/0

Table I: Physical and Translational Characteristics of Cotton Poly(A)+ mRNA[a]

stage	total RNA content[b] (NT × 10^{-10}/cell)	% poly(A)+ mRNA of total RNA: recovered[c]	% poly(A)+ mRNA of total RNA: calcd[d]	translatable mRNA[e]	mass av length (NT)
young embryo	1.8 ± 0.1 (3)	1.0 ± 0.1 (4)	1.5 ± 0.2 (2)	52 ± 7 (4)	1500 ± 50 (3)
older embryo	3.7	1.2		54	1500
mature embryo	3.7	1.0 ± 0.1 (2)[g]	1.2	71	1550
12-h seedling	3.6	1.2 ± 0.1 (2)		62 ± 3 (3)	
24-h seedling[h]	3.7[f]	1.6 ± 0.2 (6)	1.9 ± 0.1 (6)	59 ± 5 (5)	1620 ± 40 (6)

[a] Numbers in parentheses indicate the number of preparations tested. Uncertainties are expressed as one standard deviation. [b] The number of cells in cotyledons remains constant during this developmental period (Walbot & Dure, 1976). Two-thirds of the cells are large parenchymal cells and one-third are small epidermal cells (Walbot & Dure, 1976) which are assumed here to contain insignificant amounts of RNA. [c] Recovered by chromatography on poly(U)-agarose. [d] Deduced from the relative poly(A) contents [measured as described in Galau et al. (1981)] of unfractionated and recovered polyadenylated RNAs = [poly(A) content in total RNA/poly(A) content in poly(A)+ mRNA] × 100. [e] From a comparison of in vitro translational efficiency (radioactivity incorporated per microgram of RNA) of unfractionated and poly(A)− RNA = [1 − efficiency of poly(A)− RNA/efficiency of total RNA] × 100. [f] Harris & Dure (1978). [g] Harris & Dure (1978) report 0.8%. [h] Galau et al. (1981).

assayed independently. These additional data (reactions between homologous mRNAs and cDNAs from young embryos and 24-h germinated cotyledons) have been pooled in Figure 2 with the results from the first experiment described above.

DNA was prepared from cotton cotyledons as described (Walbot & Dure, 1976) and sheared to a 650-nucleotide single-strand length by sonication. Single-copy DNA was prepared by two successive incubations to C_0t = 125 and the nonreassociated DNA isolated in each case by hydroxylapatite chromatography. It was incubated to C_0t = 5000 to form double-stranded material and radiolabeled with DNA polymerase I (Galau et al., 1976, 1977). Reassociation of DNAs with unlabeled cotton DNA was in 0.5 M sodium phosphate buffer (pH 6.9), 10 mM EDTA, and 0.2% sodium dodecyl sulfate at 68 °C. Reassociation was assayed with hydroxylapatite. The values of DNA C_0t have been multiplied by a factor of 6.1 to yield the value expected in standard conditions (Britten et al., 1974). Hybridization and reassociation data were fit by computer by nonlinear least-squares procedures (Pearson et al., 1977).

Results

Characteristics of Poly(A)+ RNA from Cotton Embryos and Seedlings. We have examined the sequences present in total poly(A)+ RNA in cotyledons from several developmental stages. There is, of course, no certainty that the sequence content and sequence concentration of total poly(A)+ RNA is the same as that of total active mRNA. However, a thorough examination of sequence content and sequence concentration in total, total poly(A)+, total poly(A)−, total polysomal, polysomal poly(A)+, and polysomal poly(A)− RNAs from the cotyledons of 24-h germinated seedlings showed no gross differences in these parameters (Galau et al., 1981). The characteristics of the other stage RNAs studied here are similar to those of 24-h germinated seedlings (Table I). Furthermore, in vitro protein synthesis in the wheat germ or rabbit reticulocyte system using total, total poly(A)+, and total poly(A)− RNAs from cotyledons from all stages studied shows no abundant mRNAs in total or total poly(A)− RNA that are not abundant mRNAs in the total poly(A)+ RNA fraction (data not shown), even though a large fraction of the translatable mRNA is nonpolyadenylated (Table I). In addition, the bulk of the sequences present in total poly(A)+ RNA from all stages are transcribed from nonrepetitive sequences (Figure 1), as expected for mRNA sequences (Lewin, 1975), and the results presented here imply that most of the sequences present in total poly(A)+ RNA during this developmental period are ultimately found in polysomes at 24 h of germination, regardless of their earlier subcellular distribution. Thus, in spite of the above possible constraint, we believe the data presented give, on a gross level, a representative picture of mRNA complexity and of mRNA abundancy changes during the developmental span studied.

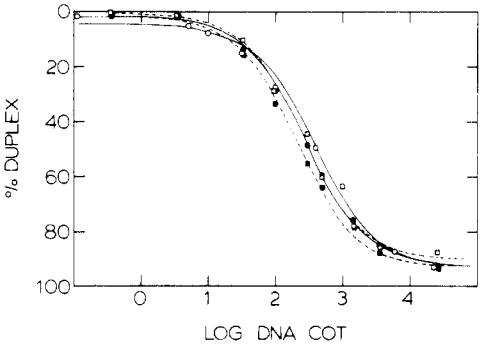

FIGURE 1: Sequence representation in young and mature embryo poly(A)+ mRNA. Eight hundred nucleotide long cDNAs synthesized from young embryo (□), mature embryo (■), and 24-h germinated seedling (○) poly(A)+ mRNA were reassociated with a 10^6-fold mass excess of 650-nucleotide young embryo cotyledon DNA. In addition, 550-nucleotide single-copy DNA was reassociated with a 1.4×10^4-fold excess of cotton DNA (●). The curves are single second-order functions fit to the data. The rate constants are 0.0038 ± 0.0004, 0.0048 ± 0.0004, and 0.0035 ± 0.0003 $M^{-1}s^{-1}$ for cDNAs made from young embryo, mature embryo, and 24-h germinated seedling mRNAs, respectively. The rate constant for the reassociation of the single-copy DNA was 0.0026 ± 0.0004 $M^{-1}s^{-1}$. Within the indicated uncertainties, all these rate constants are the same after correction for disparity in fragment lengths (Chamberlin et al., 1978).

Number of Sequences in Embryo and Seedling mRNAs. Complementary DNAs, synthesized from cotyledon poly(A)+ mRNA prepared from the three developmental stages, were each hybridized to all three mRNAs with a 3×10^3-fold mass excess ($\sim 1.5 \times 10^3$ fragment excess) of mRNA. DNA complementary to young embryo cotyledon mRNA was also hybridized with cotyledon mRNA from a stage of development intermediate between young and mature embryos (older embryo) and with mRNA from 12-h germinated seedling cotyledons. The results are presented in Figure 2. In each of the left panels the same cDNA reacts with the different mRNAs, and in each of the right panels the same data is replotted such that all three cDNAs react with the same mRNA. In order to calculate the number and concentration of mRNA sequences present in the cotyledons, the homologous cDNA–mRNA hybridization data were fit by computer with several pseudo-first-order rate components which represent the reaction of cDNA with mRNA complements present in different abundance or concentration classes (Bishop et al., 1974).

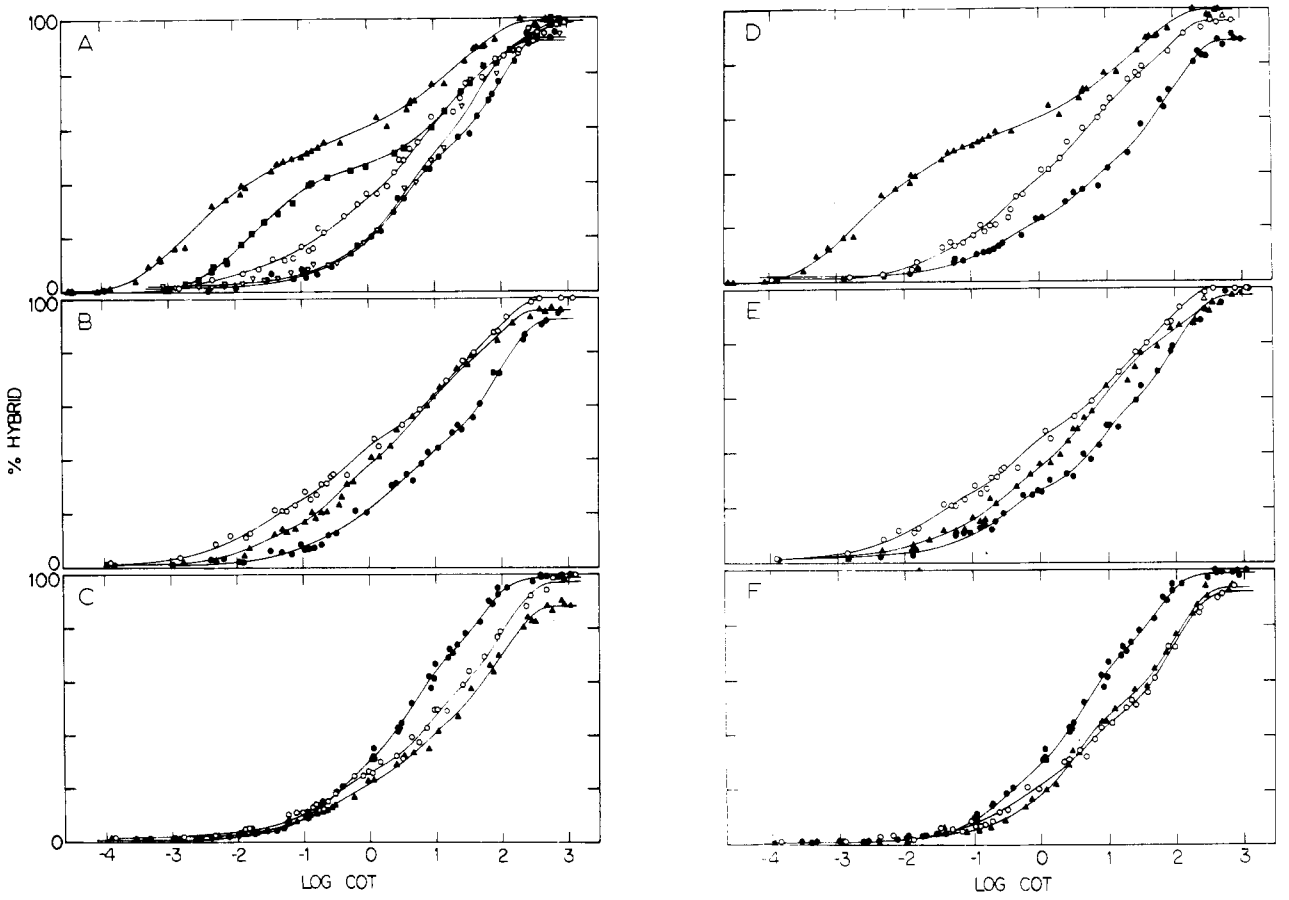

FIGURE 2: Hybridization of embryo and seedling poly(A)+ mRNAs with homologous and heterologous cDNA. Poly(A)+ mRNAs were hybridized with the cDNAs made from these mRNAs. In the left-hand panels the cDNA used in each panel is the same, the cDNA being from young embryo, mature embryo, and 24-h germinated seedling mRNA in parts A, B, and C, respectively. Poly(A)+ mRNA was from young embryos (▲), older embryos (■), mature embryos (○), 12-h germinated seedlings (▽), and 24-h germinated seedlings (●). In the right-hand panels the data are replotted such that the mRNA in each panel is the same, being mRNA from young embryos, mature embryos, and 24-h germinated seedlings in parts D, E, and F, respectively. The cDNAs were from mRNA of young embryos (▲), mature embryos (○), and 24-h germinated seedlings (●). The values of C_0t are not corrected to standard conditions due to the uncertainties in the magnitude of the correction (Van Ness et al., 1979). Similarly sized ovalbumin mRNA and cDNA react under these conditions with a rate constant of 840 ± 80 $M^{-1}s^{-1}$. The curves through the data are the results of the solutions described in the last line for each reaction in Table II.

Table II details the cDNA component sizes and rate constants calculated by using the minimum number of components which yield a minimum root mean square error (Hastie & Bishop, 1977) and which are also consistent with the reactions of these same mRNAs with heterologous cDNAs (see below). The average number of different mRNAs sequences in each component are calculated by the method of Bishop et al. (1974).

Cotyledon mRNAs from each of the three stages can be described as containing several abundance classes of mRNA. As development proceeds, there is a gradual disappearance of the two most abundant mRNA classes found in the young embryo. About 30% of the mRNA in young embryos behaves kinetically as a single mRNA sequence. This fraction, labeled component 1 in Table I, undoubtedly contains the major storage protein mRNAs. A large fraction of the in vivo and in vitro protein synthesis in this tissue is of two major molecular weight forms (each with isoelectric point variants) of storage protein (Dure & Chlan, 1981; Dure & Galau, 1981). The two forms have similar amino acid compositions and cross-react immunologically (Dure & Chlan, 1981). As predicted from the relatedness of the proteins, the rate of hybridization of component 1 shows that the various forms of storage proteins are coded for by very similar mRNAs which can cross-react under the hybridization conditions employed here. The general observation may be made that the number of cotyledon mRNA sequences in each of the other abundance classes appears to increase ~2-fold between young and mature embryo stages and then declines during the first 24 h of germination. The total number of cotyledon sequences present from this analysis is about 14000, 34000, and 21000 in young embryo, mature embryo, and 24-h germinated seedlings, respectively.

Qualitative Changes in mRNA Concentration with Development. For examination of the representation and change in concentration of mRNAs in each abundance class with development, each cDNA was hybridized with the other stage mRNAs, as previously mentioned (Figure 2). Two principle points may be made from the heterologous reactions. First, it is clear that the more abundant embryonic mRNAs exist in much lower concentrations in other stage RNAs. Secondly, at least 90% of the cDNAs react with all mRNAs. The fragment ratio of mRNA to cDNA used in these experiments, $\sim 1.5 \times 10^3$, should be sufficient, at infinite RNA C_0t, to drive cDNA into hybrid if the concentration of the mRNA complement in the heterologous mRNA is as little as $1/(1.5 \times 10^3)$ or 7×10^{-4} the level in the homologous mRNAs. These observations imply that the more abundant sequences are shared in all tissues in concentrations within 10^3 of each other and (since the hybridization reactions were taken to RNA C_0t = 1000 only) that many of the less abundant mRNAs are shared at concentrations within 10^1–10^2 of each other.

Quantitative Interpretation of Reciprocal cDNA–mRNA

Table II: Summary of cDNA-mRNA Hybridization with Embryo and Seedling Cotyledon Poly(A)+ mRNAs[a]

			component								
			1			2			3		
rxn	mRNA	cDNA	% cDNA	K ($M^{-1}\cdot s^{-1}$)	no. of mRNAs[b]	% cDNA	K ($M^{-1}\cdot s^{-1}$)	no. of mRNAs[b]	% cDNA	K ($M^{-1}\cdot s^{-1}$)	no. of mRNAs[b]
1	young embryo	young embryo	28 ± 4	540 ± 110	0.6	18 ± 4	49 ± 28	4	11 ± 3	4.5 ± 4.1	28
4		mature embryo				12 ± 2	33 ± 10		30 ± 3	1.4 ± 0.3	
						6 ± 2	78 ± 56		22 ± 2	≡4.5	
						9 ± 1	≡49		22 ± 2	≡4.5	
7		24-h seedling							22 ± 2	1.9 ± 0.3	
									14 ± 2	≡4.5	
									14 ± 2	≡4.5	
5	mature embryo	mature embryo				17 ± 2	64 ± 18	3	23 ± 2	3.5 ± 0.8	75
2		young embryo				6 ± 2	89 ± 110		19 ± 2	4.4 ± 1.6	
						6 ± 2	79 ± 76		22 ± 2	≡3.5	
8		24-h seedling							21 ± 2	4.0 ± 0.7	
									19 ± 1	≡3.5	
									19 ± 1	≡3.5	
9	24-h seedling	24-h seedling							18 ± 2	4.5 ± 0.7	46
3		young embryo							5 ± 1	71 ± 48	
									7 ± 1	≡4.5	
6		mature embryo							18 ± 3	2.6 ± 0.7	
									14 ± 2	≡4.5	
									14 ± 1	≡4.5	
10	older embryo	young embryo				23 ± 3	80 ± 14		23 ± 3	94 ± 1.6	
						34 ± 2	45 ± 6		9 ± 2	≡3.5	
11	12-h seedling	young embryo							6 ± 2	19 ± 16	
									7 ± 2	≡4.5	

[a] The hybridization reactions in Figure 2 were fit by computer as described in the text. In some cases the rate constants were constrained. The statistical uncertainties of the calculated parameters are expressed as one standard deviation. [b] The numbers of mRNAs in each abundance class in the homologous cDNA–mRNA hybridization reactions were calculated according to Bishop et al. (1974) with reference to a parallel reaction (not shown) of 1900 nucleotide long ovalbumin mRNA and 800 nucleotide long ovalbumin cDNA (K = 840 ± 80 $M^{-1}\cdot s^{-1}$).

Hybridizations. For quantitative analysis of RNA concentration changes, we have attempted to discover what fraction of the sequences in each abundance class is shared in the other mRNAs and at what concentration. That is, we have tried to follow the movement of sequences from one abundancy class to another during this developmental span. The data presented in Figure 2 illustrate a common problem in the analysis of heterologous reactions between cDNAs and different stage mRNAs. The presence of several mRNA abundance classes in each stage mRNA makes it difficult, even qualitatively, to deduce which cDNA component is reacting with complements in each heterologous mRNA abundance class. One solution to this problem has been to augment these studies with hybridizations of kinetically purified cDNAs (complementary to mRNAs in only one or a limited number of RNA abundance classes) to heterologous mRNAs so that the fraction which reacts, and the rates at which it does so, allows the calculation of the concentrations of sequences shared in the heterologous mRNA [e.g., Levy & McCarthy (1975) and Hastie & Bishop (1976)]. Due to the large number of stages we wished to examine and the large number of abundance classes in each stage mRNA, we have extended the use of reciprocal cDNA–mRNA hybridization to include all possible complementary hybridizations (nine) between the three stage mRNAs and their cDNAs (Figure 2). Applying a reasonable assumption as to the behavior of such reactions, it was found that it was possible to deduce to which heterologous mRNA abundance class each abundance class cDNA, or fraction thereof, hybridized without the use of kinetically selected cDNA fractions. The analysis of the data shown in Figure 2 and Table II is outlined below in three parts. The first step is to determine how much of the heterologous cDNA reacts with sequences in each mRNA abundance class. The second is then to deduce the partition during hybridization of each heterologous abundance class cDNA into each of the abundance classes in the other mRNAs. Finally this partition is reduced to a flow of numbers of mRNA sequences in mRNA groups in which mRNAs change together in concentration during development.

In order to determine how much heterologous cDNA reacts with each abundance class in an mRNA, we have made the assumption, from the observations of Hastie & Bishop (1976), that shared sequences in an mRNA population will be present throughout the developmental span in concentrations encompassed by at least one of the mRNA's abundance classes. That is, particular mRNAs shared with an earlier stage mRNA stay in the same abundance class or move from one abundance class to another and do not long exist in concentrations intermediate between the principal classes. This must certainly be true when there is very extensive homology between the sequences in the same abundance class in two different mRNAs. Where abundance classes contain very few sequences or where the number of shared sequences in an abundance class constitutes a minor fraction of the total number of sequences in that class, this assumption is not logically required. However, in support of this assumption, at least two studies involving the hybridization of kinetically fractionated cDNAs with dissimilar mRNAs tend to show that the rates of hybridization of heterologous cDNAs are very similar to the rates of homologous cDNA hybridization (Hastie & Bishop, 1976; Goldberg et al., 1981). Moreover, many of the heterologous reactions with cotton mRNAs are entirely consistent with this notion.

Each heterologous reaction (as presented in the right-hand panels in Figure 2) was initially fit by computer with the number of components required to yield a minimum root mean square error (Hastie & Bishop, 1976) and which was consistent with the number of abundance components in the mRNA (from the homologous cDNA–mRNA reactions). In accordance with the assumption made above, a further series of solutions was conducted. Operationally, we have assumed that

component									
4			5			6			
% cDNA	K ($M^{-1} \cdot s^{-1}$) ($\times 10$)	no. of mRNAs[b]	% cDNA	K ($M^{-1} \cdot s^{-1}$) ($\times 10^2$)	no. of mRNAs[b]	% cDNA	K ($M^{-1} \cdot s^{-1}$) ($\times 10^3$)	% cDNA NR[c]	rms[d] error (%)
15 ± 4	2.4 ± 1.3	700	26 ± 4	2.2 ± 0.4	13 000			1 ± 1	1.5
30 ± 3	1.1 ± 0.3		24 ± 2	0.86 ± 0.21				3 ± 1	1.4
32 ± 2	≡2.4		20 ± 4	≡2.2		14 ± 3	5.8 ± 0.48	4 ± 1	1.7
32 ± 2	≡2.4		30 ± 1	1.2 ± 0.2				5 ± 1	1.6
30 ± 4	0.47 ± 0.12		34 ± 4	0.58 ± 0.12				10 ± 1	1.7
18 ± 2	≡2.4		31 ± 4	≡2.2		25 ± 3	4.3 ± 1.2	10 ± 1	1.7
22 ± 2	≡2.4		50 ± 2	1.1 ± 0.1				12 ± 1	2.0
28 ± 4	1.4 ± 0.5	2 300	31 ± 4	1.1 ± 0.2	32 000			0 ± 1	1.9
43 ± 2	1.9 ± 0.3		27 ± 1	0.72 ± 0.12				4 ± 1	1.9
43 ± 2	≡1.4		26 ± 2	0.55 ± 0.11				2 ± 1	2.0
35 ± 3	0.84 ± 0.14		40 ± 3	0.61 ± 0.11				2 ± 1	1.7
27 ± 2	≡1.4		37 ± 8	≡1.2		12 ± 7	3.7 ± 2.4	2 ± 1	1.8
27 ± 2	≡1.4		49 ± 2	0.93 ± 0.04				3 ± 1	1.9
38 ± 2	3.0 ± 0.4	1 400	41 ± 2	2.3 ± 0.2	20 000			1 ± 1	1.4
38 ± 1	3.8 ± 0.3		51 ± 2	1.1 ± 0.2				6 ± 1	1.4
36 ± 2	≡3.0		49 ± 1	1.0 ± 0.1				7 ± 2	1.6
21 ± 3	2.1 ± 0.7		51 ± 2	1.0 ± 0.1				8 ± 2	1.7
24 ± 2	≡3.0		20 ± 9	≡2.3		34 ± 7	7.2 ± 2.0	7 ± 1	1.7
25 ± 2	≡3.0		52 ± 1	1.1 ± 0.1				8 ± 2	1.7
30 ± 2	0.73 ± 0.09		24 ± 2	0.51 ± 0.08				0 ± 1	1.0
29 ± 1	≡1.4		26 ± 1	0.67 ± 0.10				1 ± 1	1.7
36 ± 3	3.0 ± 0.6		48 ± 3	1.9 ± 0.3				8 ± 1	1.8
36 ± 3	≡3.0		48 ± 2	2.0 ± 0.2				8 ± 1	1.9

The cDNA in these reactions averaged 800 nucleotides in length, and we assume a number average length of cotton mRNAs of 1400 nucleotides (Galau et al., 1981). Thus the number of mRNAs = (840/K)(1900/1400)/(% of cDNA in component). The maximum possible correction for rate differences in the ovalbumin and cotton mRNA hybridization reactions due to disparity in the mRNA lengths (Chamberlin et al., 1978) would increase the numbers of mRNAs by a factor of <1.15. [c] NR = not reacted. [d] rms = root mean square.

the rate constants with which a particular mRNA hybridizes with any heterologous cDNA will be a function of the mRNA and only the fraction of cDNA reacting with each abundance class mRNA will vary. Thus the pseudo-first-order functions were fit to the data constraining at least some of the rate constants in the reactions of a particular mRNA with heterologous cDNAs to those describing the reaction of that mRNA with its homologous cDNA. The results of some of these solutions are presented in Table II (reactions 2–4 and 6–8).

The hybridization data are equally well fit if all rate constants are left free or if all the rate constants are constrained for components 2–4. With one exception (reaction 3, component 3), there is fairly good agreement in the rate constants for these components for the same RNA hybridizing with all cDNAs. The average ratios of nonrestrained heterologous rate constants to the rate constants for the homologous cDNA reactions are 1.2 ± 0.65 (N = 2), 0.91 ± 0.22 (N = 5), and 0.77 ± 0.23 (N = 6) for components 2–4, respectively. In addition, the best estimates of the component sizes in the nonconstrained fits are very close to those calculated by assuming the homologous rate constants. The average ratios of the component sizes in the nonconstrained fits to those in the constrained fits are 0.73 ± 0.38 (N = 2), 1.11 ± 0.11 (N = 6), and 1.30 ± 0.14 (N = 6) for components 2–4, respectively. These observations provide strong support for the simplifying assumption made above. In all but one of the heterologous reactions, the hybridization of the most slowly hybridizing cDNA could be fairly well described as a single component, but with an average rate constant of 0.46 ± 0.04 (N = 5) relative to that of the homologous reaction. When the rate constants for the more abundant components were constrained, the nonconstrained rate constant for component 5 was still only 0.50 ± 0.02 (N = 6) of the homologous cDNA–mRNA reaction for component 5. Attempts to constrain all rate constants, including that of component 5, led to much higher root mean square errors in the majority of the solutions. A systematic investigation was conducted to evaluate whether the hybridization of the heterologous cDNAs could be described as consisting of two components, one with a rate constant similar to that of component 5 in the homologous cDNA–mRNA reaction and one significantly slower (putative component 6). In two reactions (2 and 3) no component 6 was detected; thus, the heterologous rate constant for component 5 must be ~0.5 that at which the mRNA drives its homologous cDNA. In the remaining four reactions a component 6 could be fit to the data, but only in reactions 4 and 7 were the solutions at all convincing. The large predicted differences in the numbers of mRNAs in component 5 in young and mature embryos (Table II, reactions 1 and 5) and the extensive cross-reaction of the cDNAs would predict the presence of a component 6 in reactions 4 and 7. We conclude that the data in Figure 2 alone do not allow rejection of either alternative set of solutions. Whichever solutions are used, the quantity of heterologous cDNA reacting with the more abundant mRNA classes (2–4) is affected *very* little. This is true for constrained solutions with or without component 6 in reactions 4 and 7 or for reasonable interpretations of most of the nonconstrained solutions. In the analysis which follows, we have examined the consequences of both alternative descriptions of these reactions with the rate constants constrained for the more abundant components. We label the solution without component 6 (last entry for each reaction in Table II) as alternative A and that with component 6 in reactions 4 and 7 as alternative B. For simplicity we have restricted the presentation of the analysis of alterative B to only that of the lowest abundance classes.

In these studies we also wished to determine the flow of abundant young embryo mRNAs in the mRNAs of develop-

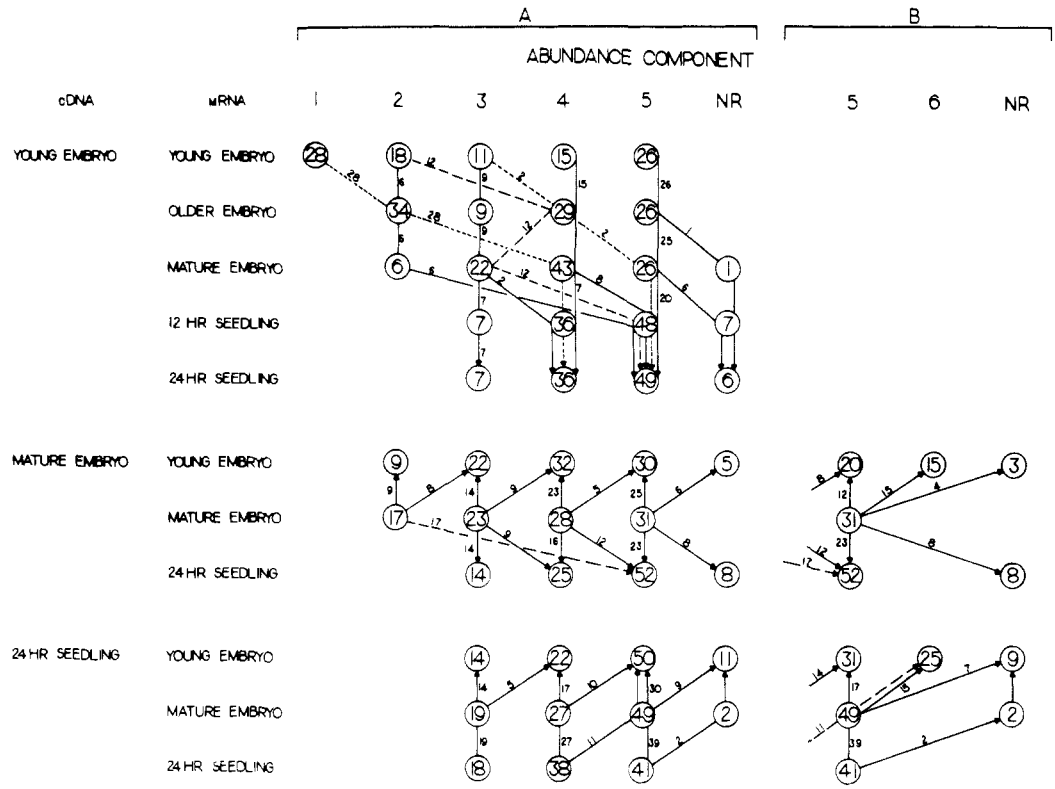

FIGURE 3: Partition of abundance class cDNAs in heterologous mRNA abundance classes. The individual matrices represent the percent of nonfractionated cDNA which reacts with each mRNA abundance class in each of the mRNAs. These numbers are the results of the solutions described in Table II. The arrows, and their associated numbers (as a percent of the total cDNA), show the portion of each abundance class cDNA which reacts with mRNAs in each of the heterologous mRNA abundance classes. A and B refer to the results obtained with alternative A and alternative B, respectively. NR = nonreacted.

mental stages intermediate to those which were systematically investigated above. Reactions 10 and 11 are of cDNA made from cotyledon mRNA from young embryos hybridized with the mRNA from cotyledons of older embryos and 12-h germinated seedlings (Table II; Figure 2A). A thorough investigation would require at least the reactions of these mRNAs with homologous cDNA, ideally a total of 16 reactions to describe the concentration of sequences shared in all other stage mRNAs. Due to the similarity of the rate constants for the abundant components in the hybridization reactions with the other three stage mRNAs, these data have been fit with the hybridization rate constants achieved with developmentally related stage mRNAs. Within these limits, the results are essentially independent of the values assumed for the rate constants for components 3 and 4.

The next step was to determine what fraction of each abundance class cDNA reacts with each mRNA abundance class in the other mRNAs. The amount of unfractionated cDNA which reacts with each abundance class mRNA in the three stage mRNAs is shown in Figure 3. The most probable partition of the sequences in each abundance class in the homologous mRNA into heterologous mRNA abundance classes is indicated by arrows. In the construction of this figure, the hybridization of the cDNA complementary to each of the abundance classes in the homologous mRNA was chosen which was consistent with all reactions, with the RNA to cDNA ratio, and with the relative number of sequences estimated (from the homologous reactions) to be in each abundance class in the three mRNAs. For the logic involved and a discussion of the method, see Appendix. If the amounts of cDNA reacting with each mRNA abundance class are assumed to be as shown, there are very few instances where an alternative partition is possible. These alternatives involve minor fractions of the mRNAs and do not qualitatively affect the overall hybridization pattern. It should be noted that only the bulk of the cDNA can be traced in each set of reactions. Minor fractions may be distributed differently.

From this analysis the temporal change in concentration of several major groups of mRNAs may be calculated (Figure 4). In this last step the numbers of sequences in each abundance class were calculated by using each of the estimates derived from the homologous reactions in Table II and the partition of cDNAs in the heterologous reactions as shown in Figure 3. Analysis of all possible reciprocal reactions leads to powerful logical constraints which tend to minimize the importance of otherwise large uncertainties in the calculated fraction of cDNA which reacts with individual abundance class mRNAs. We are unable, however, to clearly identify either alternative A or B as being more reasonable. Certainly alternative B best agrees with the apparent complexities of the mRNAs as calculated from the homologous cDNA–mRNA reactions. In addition, a component 6, comprising the required 5% of the total cDNA, can easily be accommodated in the young embryo mRNA reaction with its homologous cDNA. Further experiments are required to make firm conclusions as to the total number of cotyledon mRNAs and their representation in component 5.

Discussion

Reciprocal Heterologous cDNA–mRNA Hybridization. The method we have used is an application of complementary or reciprocal cDNA–mRNA hybridization reactions [e.g., Ryffel & McCarthy (1975), Levy & McCarthy (1975), Young et al. (1976), Axel et al. (1976) and Hastie & Bishop (1976)] and follows from the observation of Hastie & Bishop (1976) that mRNAs tend to hybridize dissimilar heterologous cDNAs at predictable rates. mRNAs from three stages of cotyledon development were studied here, with nine hybrid-

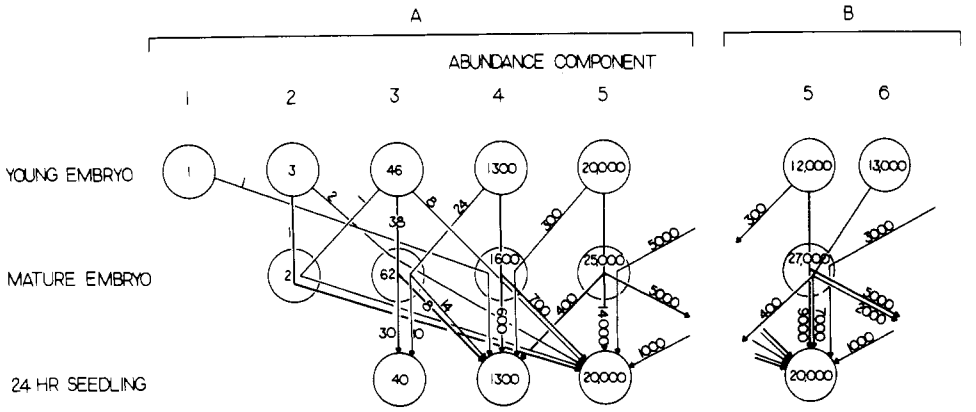

FIGURE 4: Changes in concentration of major mRNA subsets with development. The mRNA abundance classes are indicated by the circles. The number within is the total number of different mRNA sequences calculated to be in each class. Individual groups of mRNAs are indicated by the arrows. The associated numbers are the numbers of different mRNA sequences in each group. A and B refer to the results obtained with alternative A and alternative B, respectively.

ization reactions. These data proved sufficient to provide an unambiguous assignment of mRNA changes in all but the least abundant mRNA class. The particular mRNAs studied probably provided a stringent test of the method of analysis. The number of reactions required increases with the square of the number of stages studied, but it was observed that only a few selected reactions with additional mRNAs provided important detailed information when analyzed in context with the reciprocal hybridization data with the other mRNAs. We believe that the method of analysis presented here may be of use in interpreting heterologous hybridization reactions with as few as two mRNA populations or from mRNA populations from tissues which are not temporally related, as well as in reactions with kinetically fractionated cDNAs. It is anticipated that uncertainties encountered in such a study might be resolved with selected use of kinetically fractionated cDNAs in cDNA–mRNA hybridizations to precisely determine some of the rate constants in homologous and heterologous mRNA reactions or with independent measurements of the number of sequences in the least abundant mRNA classes by single-copy DNA saturation hybridization (Galau et al., 1974). Such studies would be necessary in the present case to resolve the described uncertainty in the distribution of mRNAs in the least abundant mRNA class in cotton cotyledons. In addition, variation in the mRNA to DNA ratio in the hybridization reactions should, in at least some cases, resolve potential ambiguities.

Although not fully exploited here, it appears, as might be predicted, that in some instances the rate constants obtained in heterologous cDNA–mRNA reactions could be of use in evaluating the best estimate of the rate constants in the homologous reactions. If the rate constants are as predictable as they appear to be, it follows that the most appropriate evaluation of these sorts of data might involve simultaneously finding the rate constants which best describe all the reactions with a particular mRNA.

Sequence Changes in mRNAs during Cotton Cotyledon Development. The approach we have taken in evaluating the sequence changes during cotyledon development has provided a strikingly detailed picture of large-scale changes in the concentration of many mRNAs. A minimum of 17 groups of mRNAs is observed. We define a subset as a group or groups of mRNAs which change with time in a unique fashion. One observes a minimum of 11 mRNA subsets, at least three of which contain two or three groups, each in a different mRNA abundance class. We consider this a minimum estimate. Putative groups of mRNAs containing a small fraction of the sequences in an abundance class are apt to be missed. In addition, we have examined in detail only three time points in this tissue's ontogeny. It is clear from Figure 3 that even a cursory examination of the intermediate developmental stages detects additional mRNA groups.

We have observed at the mRNA sequence level the same abundant mRNA subsets as were observed by the analysis of in vivo and in vitro protein synthesis in this tissue during the same period [subsets numbered 1–5 in Dure et al. (1981, preceding paper)]. Several hundred proteins were detected in those studies, though only ~80 were very abundant. In agreement with this observation, only 50–70 proteins are predicted to be abundant from the hybridization reactions here (components 1–3). As presented in Figure 4, there are sequences in components 1–3 which change very little in concentration throughout this developmental time span (subset 1), those which decline in concentration only in early germination (subset 2), and some which decline in late embryogenesis (subset 3). There are, in addition, sequences which increase in concentration in the late stages of maturation, some of which remain abundant in the first hours of germination (subset 4), and the rest of which decline during early germination (subset 5). The heterologous hybridization given in reaction 8 of Table II did not detect the few (about six) abundant mRNAs which are possibly unique to germination (Dure et al., 1981), though the data do indicate that many less abundant mRNAs have increased severalfold in their concentration during the first 24 h of germination. In general, the numbers of mRNAs predicted in each of these subsets from the hybridization data are very similar to those observed by examination of the in vivo protein synthesis.

The nucleic acid hybridization experiments reported here extend the observations of Dure et al. (1981) in one very important aspect. Stage-specific changes in abundant proteins and their mRNAs occur in parallel with similar changes in mRNAs present in lower concentrations. Thus it is likely, within the resolution afforded by these experiments, that we have evidence that whatever signals affect the steady-state level of mRNA subsets, affect groups of mRNAs in a similar fashion regardless of their relative concentration. If this is so, there must exist several independent levels at which mRNA synthesis and turnover are specifically regulated.

Especially striking in Figure 4 is the large reduction in the concentration of the storage protein mRNA and of several mRNAs in component 2 in young and mature embryos during the course of development studied here. The storage protein mRNA declines ~10-fold in the older embryo (Figure 3) and

Table III: Number of mRNAs Present in Higher Plants

plant and tissue	RNA	no. of mRNA species for each abundance class (% of total mRNA in each class)					
		kinetics of cDNA hybridization[a]					single-copy DNA saturation[b]
		1	2	3	4	5	5
tobacco							
leaf[c]	polysomal						27 000
	polysomal A+		10 (9)	340 (52)		11 000 (39)	
five others[d]	polysomal						26 000
parsley[e]							
root callus	polysomal						13 000
	polysomal A+				960 (56)	11 000 (44)	13 000
leaf	polysomal						10 000
	polysomal A+		28 (29)			9 200 (70)	
barley shoot[f]							
green	polysomal A+		14 (31)	270 (32)	1 200 (6)	31 000 (31)	
etiolated	polysomal A+		12 (15)		2 000 (28)	31 000 (52)	
soybean							
suspension cells[g]	A+			57 (18)	1 900 (44)	30 000 (38)	23 000
root[h]	polysomal A+			50 (19)	2 000 (36)	18 000 (45)	
nodule[h]	polysomal A+	1 (24)			510 (25)	22 000 (51)	
hypocotyl[i]	A+		2 (12)	200 (30)	2 300 (24)	32 000 (34)	44 000
plus auxin[i]	A+			67 (21)	2 200 (38)	38 000 (41)	44 000
leaf[j]	polysomal A+		2 (8)	45 (17)	900 (35)	35 000 (22)	
embryo[j]							
very young	polysomal						14 000
	polysomal A+			180 (36)		14 000 (62)	
midmaturation	polysomal A+	1 (20)	6 (31)	180 (27)		32 000 (22)	
axis	polysomal						18 000
cotyledon	polysomal						15 000
cotton cotyledon[k]							
embryo							
young	A+	1 (28)	3 (18)	37 (11)	1 000 (15)	16 000 (26)	
mature	A+		2 (17)	68 (23)	2 000 (28)	28 000 (31)	
24-h seedling	A+[l]			43 (18)	1 400 (38)	20 000 (41)	

[a] A parallel kinetic standard was used in the calculation (Bishop et al., 1974) of the number of mRNAs. [b] Several different methods of calculating the numbers of genes from the saturation have been employed. [c] Goldberg et al. (1978). [d] Kamalay & Goldberg (1980). [e] Kiper et al. (1979). [f] Heinze et al. (1980). [g] Silflow et al. (1979). [h] Auger et al. (1979). [i] D. Baulcomb and J. Key, personal communication. [j] Goldberg et al. (1981). [k] This work. The quantity of mRNA in each component is taken from Table II. The numbers of genes are the averages of the estimates in Table II and in Figure 4 using alternative A. [l] All RNA fractions studied had the same sequence composition (Galau et al., 1981).

an additional 100-fold in mature embryos. Its concentration appears not to change very much during the first 24 h of germination. This observation is entirely consistent with the synthesis of this protein in vivo (Dure & Chlan, 1981). At least with respect to sequences which are shared with young embryo mRNA and decline in concentration in the first 24 h of germination, all of these changes occur in the first 12 h of germination (Figure 3). Similar declines are noted in some of the RNAs in many of the abundance classes; for instance, 2 sequences in component 2 in mature embryos and about 5000–7000 sequences in component 5 in mature embryos decline at least 100-fold in concentration in the first 12 h of germination.

Our observations in cotton are in general agreement with a recent study of the mRNAs present during soybean embryogenesis (Goldberg et al., 1981). In that organism there are ~10 polysomal mRNA species which comprise together ~50% of the polysomal poly(A)+ mRNA mass during midembryogenesis. These sequences decline precipitously in concentration in late embryogenesis but are still detectable in mature embryo total RNA. Most of the embryo sequences are present in polysomes at two stages of embryogenesis and in mature embryo RNA, but, as is observed here in cotton, many vary significantly in concentration in these tissues.

The numbers of mRNAs and their concentration we have measured in cotton cotyledons are similar to those described in other plant tissues. A compilation of the results of studies so far published is presented in Table III. There exists a wide range of values for the total number of active genes, ranging from 9000 to 44 000. Some of this variation probably reflects the experimental method and differences in interpretation of the complexity of nonrepetitive DNAs. In our own experiments, we are unable at present to define precisely the number of active genes in cotyledons. Several observations may be made however from the studies listed in Table III. There are several instances of superabundant mRNAs in plant tissues; these include mRNAs for leghemoglobin (Auger et al., 1979), storage proteins, and other seed-specific proteins (Goldberg et al., 1981; this study), as well as several in leaves (Goldberg et al., 1981). Large-scale changes occur in the concentration of many abundant mRNAs and less abundant mRNAs as well between tissues and within a tissue with development (Auger et al., 1979; Kamalay & Goldberg, 1980; Baulcomb & Key, 1980; Heinze et al., 1980; this study).

Appendix

An analysis of heterologous hybridization reactions yielded the fraction of the total cDNA which reacted with each heterologous mRNA abundance class. The data were arranged as shown in Figure 3 in the text into three matrices, each matrix representing the reaction of a given cDNA with the homologous and heterologous mRNAs. The object was to then deduce which portion of the cDNA hybridizes with complements in each of the heterologous mRNA abundance classes. Here we outline the rationale used in this analysis and present some examples to illustrate the method.

Several general statements concerning the hybridization may be made. (1) The cDNAs must either react with an abundance component in heterologous mRNA or remain nonreacted. The mRNA to cDNA fragment ratio in this experiment was $\sim 10^3$; cDNAs may react with complements in heterologous mRNAs which are present in $\sim 10^{-3}$ the concentration in the homologous RNA, which here is three abundance components removed. (2) We have assumed that cDNAs will react with heterologous mRNAs in the same abundance class as their complements in the homologous mRNA or in a lower abundance class. Due to the approximately 10–20-fold difference in the number of sequences in each abundance class, only a small fraction of the mass of an abundance class' cDNA may have complements at higher concentrations in heterologous mRNAs. We have also assumed that mRNAs do not decrease and then increase in concentration with time. Thus the amount of cDNA hybridizing with an mRNA abundance class from a tissue two stages removed cannot be greater than the amount which hybridizes with that abundance component in the intermediate stage mRNA. There was one notable exception to this rule, but we find no evidence of there being others. (3) When the cDNA represents a very small number of sequences, such as in components 1 and 2 in the mRNAs studied here, a large fraction of the total cDNA in the homologous mRNA abundance class must hybridize elesewhere as a whole, since it represents a single sequence. (4) The hybridization pattern in each matrix must be consistent with that in the other two. As an example, young embryo mRNAs in a particular abundance class, which decline in concentration between the mature embryo and 24 h of germination, must be a subset of all mature embryo mRNAs in that abundance class which also decline in concentration during the 24-h germination. The young embryo and mature embryo cDNA complements, or fractions thereof, thus must hybridize with the same abundance class mRNAs in 24-h germinated seedling mRNA. (5) In order to describe the hybridization in components 4 and 5, it was necessary to assume the relative complexities, as calculated in the homologous cDNA–mRNA reactions (Table II), of the components in some of the mRNAs. If these were unknown, or believed to be unreliable, consistent solutions could be made in some, but not in all, alternatives.

The analysis was started with assignments in the most abundant classes, working toward the least abundant classes within a single matrix. When an ambiguity was encountered, the other matrices were explored in turn until it was necessary to consider the relationship between the matrices. When the distribution of cDNA hybridization within one abundance component in all three matrices was solved, the number of sequences in the abundance components was calculated from the fraction of the sequences held in common and of the number of sequences, which is derived from the homologous cDNA–mRNA reactions. The numbers of mRNAs in each of the groups which change in concentration could then be calculated. The next lowest abundance class was investigated in a similar manner. Where a small fraction of the cDNA was involved, provisional assignments were made until the analysis was further advanced. In no case did this significantly alter the other solutions.

Some specific examples illustrate the method. (1) For component 1 in young embryo mRNA, independent data indicate that this is a single mRNA species, the principal storage protein mRNA. All 28% of the cDNA must hybridize together. It cannot remain nonreacted with the other mRNAs, since the nonreacted fraction is much too small, and can thus only hybridize to older embryo mRNA component 2 and mature embryo component 4 without violating the RNA to DNA ratio restriction. (2) Component 2 in young and mature embryos is estimated to contain about three to four sequences from the homologous hybridization kinetics. About 6% out of the 18% of the component 2 cDNA in young embryos must hybridize with component 2 in older embryos to account for the total hybridization there and thus remain in component 2 in mature embryo mRNA. The reactions with mature embryo cDNA in the middle matrix indicate that there are probably three sequences in component 2 in young embryos mRNA and two sequences in component 2 in mature embryo mRNA, one of which is shared by both stages. Further assignment of the 6% of young embryo cDNA hybridizing with component 2 in mature embryo mRNA was delayed until it was recognized that components 3 and 4 could be solved without requiring its hybridization with either of them, and thus it had to hybridize with component 5 or remain unreacted in seedling mRNAs. About 9% out of the 17% of the mature embryo cDNA in component 2 must be the same sequence as the 6% of young embryo component 2 cDNA. This sequence, as well as the second sequence in mature embryo component 2, had to hybridize with mRNAs in component 5 in seedling mRNA in order to obtain a reasonable solution of component 5.

The remaining 12% of the original 18% of the young embryo cDNA in component 2 can hybridize with older embryo mRNA only with component 4 mRNAs since (a) most of the 11% of the young embryo component 3 cDNA has to hybridize with component 3 mRNAs in the mature embryo (deduced in middle and bottom matrices), (b) component 5 is already accounted for, and (c) there is not any unreacted cDNA in the reaction of young embryo cDNAs with older embryo mRNAs. Thus, it has hybridized with component 4 in older embryo mRNA. It then must hybridize with the mRNAs in component 3 in mature embryo mRNA, since no other source of cDNA in available to account for the total hybridization of cDNA with this mRNA. The subsequent hybridization of the two mRNA sequences in this 12% of the young embryo cDNA with component 5 mRNAs in seedling mRNAs is consistent with the indicated solutions for component 5.

References

Auger, S., Baulcomb, D., & Verma, D. P. S. (1979) *Biochim. Biophys. Acta 563*, 496–507.

Axel, R., Feigelson, P., & Schutz, G. (1976) *Cell (Cambridge, Mass.) 7*, 247–254.

Baulcomb, D. C., & Key, J. L. (1980) *J. Biol. Chem. 255*, 8907–8913.

Bishop, J. D., Morton, J. G., Rosbash, M., & Richardson, M. (1974) *Nature (London) 250*, 199–204.

Britten, R. J., Graham, D. E., & Neufeld, B. R. (1974) *Methods Enzymol. 29E*, 363–418.

Chamberlin, M. E. Galau, G. A., Britten, R. J., & Davidson, E. H. (1978) *Nucleic Acids Res. 5*, 2073–2094.

Dure, L., III, & Chlan, C. (1981) *Plant Physiol.* (in press).

Dure, L., III, & Galau, G. A. (1981) *Plant Physiol.* (in press).

Dure, L., III, Greenway, S. C., & Galau, G. A. (1981) *Biochemistry* (preceding paper in this issue).

Galau, G. A., Britten, R. J., & Davidson, E. H. (1974) *Cell (Cambridge, Mass.) 2*, 9–21.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., & Davidson, E. H. (1976) *Cell (Cambridge, Mass.) 7*, 487–503.

Galau, G. A., Lipson, E. D., Britten, R. J., & Davidson, E. H. (1977) *Cell (Cambridge, Mass.) 10*, 415–432.

Galau, G. A., Legocki, A. B., Greenway, S. C., & Dure, L., III (1981) *J. Biol. Chem. 256*, 2551–2560.
Goldberg, R. B., Hoschek, G., Kamalay, J. C., & Timberlake, W. E. (1978) *Cell (Cambridge, Mass.) 14*, 123–131.
Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S., & Breidenbach, R. W. (1981) *Dev. Biol.* (in press).
Harris, B., & Dure, L., III (1978) *Biochemistry 17*, 3250–3256.
Hastie, N. D., & Bishop, J. O. (1976) *Cell (Cambridge, Mass.) 9*, 761–774.
Heinze, H., Herzfeld, F., & Kiper, M. (1980) *Eur. J. Biochem. 111*, 137–144.
Kamalay, J. C., & Goldberg, R. B. (1980) *Cell (Cambridge, Mass.) 19*, 934–946.
Kiper, M., Bartels, D., Herzfeld, I., & Richter, G. (1979) *Nucleic Acids Res. 6*, 1961–1978.
Levy, W. B., & McCarthy, B. J. (1975) *Biochemistry 14*, 2440–2446.
Lewin, B. (1975) *Cell (Cambridge, Mass.) 4*, 77–93.
Maxwell, I. H., Van Ness, J., & Hahn, W. E. (1978) *Nucleic Acids Res. 5*, 2033–2038.
O'Farrell, P. H. (1975) *J. Biol. Chem. 250*, 4007–4021.
Pearson, W. R., Davidson, E. H., & Britten, R. J. (1977) *Nucleic Acids Res. 4*, 1727–1737.
Ryffel, G. U., & McCarthy, B. J. (1975) *Biochemistry 14*, 1379–1385.
Salzberg, S., Levi, Z., Aboud, M., & Goldberger, A. (1977) *Biochemistry 16*, 25–29.
Silflow, C. D., Hammett, J. R., & Key, J. L. (1979) *Biochemistry 18*, 2723–2731.
Van Ness, J., Maxwell, I. H., & Hahn, W. E. (1979) *Cell (Cambridge, Mass.) 18*, 1341–1349.
Walbot, V., & Dure, L. S., III (1976) *J. Mol. Biol. 101*, 503–536.
Young, B. D., Birnie, G. D., & Paul, J. (1976) *Biochemistry 15*, 2823–2828.

In Vitro Synthesis of the Respiratory NADH Dehydrogenase of *Escherichia coli*. Role of UUG as Initiation Codon†

Maria I. Poulis, Denis C. Shaw, Hugh D. Campbell, and Ian G. Young*

ABSTRACT: The respiratory NADH dehydrogenase of *Escherichia coli* has been synthesized in vitro in a coupled transcription–translation system with cloned deoxyribonucleic acid (DNA) as template. The identity of the protein produced was confirmed by paper chromatography and electrophoresis of tryptic peptides. [^{35}S]Methionine-labeled tryptic peptides from the in vitro product were shown to comigrate with authentic methionine-containing tryptic peptides from the purified enzyme. Using a transcription–translation system derived from an *ndh* mutant, it was shown that the enzyme produced in vitro was incorporated into membrane vesicles of the mutant to give functional, cyanide-sensitive NADH oxidase activity. Radiochemical N-terminal sequencing of the synthesized NADH dehydrogenase showed that the product was a mixture of three different species, with *N*-formylmethionine, methionine, or threonine at the N terminus. The results indicated that only partial N-terminal processing was occurring in vitro and that the first residue of the unprocessed NADH dehydrogenase is *N*-formylmethionine. Since DNA sequencing has shown that this residue is encoded by UUG [Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A., & Shaw, D. C. (1981) *Eur. J. Biochem.* (in press)], this work verifies the role of UUG as a normal initiation codon.

The respiratory NADH dehydrogenase of *Escherichia coli* is located together with the other components of the electron-transport chain in the inner or cytoplasmic membrane of this organism. It plays a central role in energy metabolism since it catalyzes the transfer of reducing equivalents generated by the major catabolic pathways to the membrane-bound energy-conserving system. The enzyme is a relatively minor, though highly active, component of the cytoplasmic membrane of *E. coli*, and this has made its purification difficult [see Jaworowski et al. (1981a)]. These difficulties have recently been overcome by the cloning of the *ndh*[1] structural gene and the amplification of the enzyme levels in vivo (Young et al., 1978). The enzyme has subsequently been purified to homogeneity and characterized (Jaworowski et al., 1981a,b).

We wished to produce the enzyme in vitro with the cloned DNA as template, in order to facilitate investigations into the regulation of synthesis of the NADH dehydrogenase, its mode of incorporation into the cytoplasmic membrane, and the structures of the unprocessed enzyme and any partially processed intermediates. The latter aspect has acquired particular significance because the determination of the sequence of the *ndh* gene and associated studies (Young et al., 1981) have suggested that a UUG codon, which normally specifies leucine, is acting as the translational initiation codon in this case. The initiating amino acid is efficiently removed in vivo by post-translational processing (Young et al., 1981).

In the present work, we describe the use of a coupled transcription–translation system to produce active, membrane-associated enzyme in vitro. The identity of the product

† From the Departments of Biochemistry (M.I.P., H.D.C., and I.G.Y.) and Physical Biochemistry (D.C.S.), John Curtin School of Medical Research, Australian National University, Canberra City, A.C.T. 2601, Australia. *Received October 23, 1980.*

[1] Abbreviations used: kbp, kilobase pairs; *ndh*, structural gene for NADH dehydrogenase; $NaDodSO_4$, sodium dodecyl sulfate; Tes, *N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid; ubiquinone-*n*, ubiquinone isoprenologue containing *n* isoprene units in the side chain; fluorescamine, 4-phenylspiro[furan-2(3*H*)-1′-phthalan]-3,3′-dione; TPCK, L-1-(*p*-toluenesulfonyl)amido-2-phenylethyl chloromethyl ketone.

0006-2960/81/0420-4178$01.25/0