Alcohol dehydrogenase gene of *Drosophila melanogaster*: Relationship of intervening sequences to functional domains in the protein

(recombinant DNA/gene isolation/pBR322/evolutionary homology)

CHEEPTIP BENYAJATI*, ALLEN R. PLACE†, DENNIS A. POWERS, AND WILLIAM SOFER‡§

Department of Biology and McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, Maryland 21218

Communicated by Hamilton O. Smith, January 15, 1981

ABSTRACT The gene that codes for *Drosophila* alcohol dehydrogenase (ADH; alcohol:NAD⁺ oxidoreductase, EC 1.1.1.1) was identified in a bacteriophage λ library of genomic *Drosophila* DNA by using ADH cDNA cloned DNA as a probe. The DNA sequence of the protein encoding region was shown to be in agreement with the amino acid sequence of the ADH. Two intervening DNA sequences (introns) were identified within the protein encoding region: one was 65 nucleotides and located between the codons for amino acid residues 32 and 33, and one was 70 nucleotides and located between the codons for amino acid residues 167 and 168. Both contained the 5' G-T and 3' A-G dinucleotides characteristic of intron boundaries of eukaryotic genes. On the basis of secondary structure predictions, the first 140 amino acid residues of *Drosophila* ADH are in an alternating β -sheet/ α -helix arrangement which is characteristic of the coenzyme binding domain of dehydrogenases. The smaller of the two introns interrupts the domain predicted to bind the adenine portion of the coenzyme.

The molecular architecture of many eukaryotic genes resembles patchwork (2-7). Interspersed within the protein encoding sequences are untranslated regions that are excised during mRNA processing (4, 6). The exact function of these intervening sequences, called "introns," is unclear (2, 6). A intriguing hypothesis is that the introns divide the coding sequence into units that represent functional domains of a protein (2-3, 7). This idea is best illustrated in the immunoglobulin genes (8-10). Such units could be used again and again in various genes to generate new protein structures (3).

The recent cloning and determination of sequence of the alcohol dehydrogenase (ADH; alcohol:NAD⁺ oxidoreductase, EC 1.1.1.1) gene of *Drosophila melanogaster* (11) provided the opportunity to test this hypothesis with a protein from a class of enzymes with well-demarcated functional domains, the dehydrogenases. All the dehydrogenases studied to date consist of at least two domains (12): one binds the coenzyme, and the other binds the substrate. The latter domain also determines substrate specificity and provides many of the residues required for catalysis. Although there is little amino acid sequence homology among dehydrogenases in the coenzyme-binding region, the secondary structure of this domain has been highly conserved (12). This has given rise to the hypothesis that the different dehydrogenases evolved by gene fusion of an ancestral nucleotide sequence coding for the coenzyme-binding domain with different nucleotide sequences coding for different catalytic properties (13).

We present here the entire nucleotide sequence of the coding region for the *Drosophila* ADH gene. Two introns have been identified, one of which is located in the midst of the region

encoding the presumptive coenzyme-binding domain. These findings are discussed in light of the proposed role of introns in eukaryotic genes.

MATERIALS AND METHODS

Materials. *Drosophila* ADH cDNA (pADH.cDNA) was cloned in pBR322 and propagated in *Escherichia coli* strain HB101 (11). The *Drosophila* bacteriophage Charon 4 λ library was that constructed by Maniatis *et al.* (14).

Restriction Analysis and DNA Sequence Determination. Restriction fragments were analyzed by electrophoresis on agarose and acrylamide gels. 5'-Terminal labeling of restriction fragments and DNA sequence analysis were performed according to Maxam and Gilbert (15). Where possible, the sequence of both strands was determined. Each restriction site was confirmed by analysis across the cleavage point. For stretches of DNA in which only one strand could be sequenced, the experiments were repeated several times.

Containment. All recombinant DNA experiments were performed in P2 facilities in accordance with National Institutes of Health guidelines.

Peptide Sequencing. The isolation, purification, and sequence determination procedures used to analyze the tryptic peptides of *Drosophila* ADH are published elsewhere (16, 17).

Secondary Structure Prediction. The computerized method of Chou and Fasman (18) was used to predict the secondary structure of the *Drosophila* ADH.

Circular Dichroism (CD) Measurements. CD spectra were measured on a Cary 60 spectropolarimeter under nitrogen flush at 25°C in 0.1 M sodium phosphate buffer (pH 7.50). The instrument had been calibrated with *d*-10-camphorsulfonic acid as specified in the operation handbook of the unit. The CD data were expressed in terms of mean residue ellipticity deg·cm²·dmol⁻¹. A mean residue weight of 107 was used. Protein concentrations were determined by amino acid analysis and absorbance at 280 nm. The percentage helix was determined as described by Siegel *et al.* (19).

Sequence Comparisons. The computer programs ALIGN, RELATE, and SEARCH and the data base available at the National Biomedical Research Foundation were used to determine the relationship among the protein sequences of *Drosophila*, yeast, and horse liver ADHs. The scoring matrix used

Abbreviations: ADH, alcohol dehydrogenase; CD, circular dichroism; kb, kilobase(s).

* Present address: Cancer Biology Program, Frederick Cancer Research Center, P. O. Box B, Frederick, MD 21701.

† Present address: Department of Biology, University of Pennsylvania, Philadelphia, PA 19104.

‡ Present address: Waksman Institute, Rutgers University, Piscataway NJ 08854.

§ To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

FIG. 1. Restriction map and strategy for determining the sequence around the ADH^S gene of *D. melanogaster*. DNA for sequencing was derived from a 4.8-kb *EcoRI* fragment of a 15-kb insert of *Drosophila* DNA cloned in the λ vector Charon 4 (14). (A) Restriction map of fragment. (B) Restriction fragments used in the sequence determination. (C) Regions covered by the sequence determination, as indicated by arrows. The region of the genomic DNA corresponding to the cDNA clone [pADH.cDNA (11)] is shown as a striped bar. (D) Coding region of two introns in the 4.8-kb *EcoRI* fragment.

was the mutation data matrix at 250 PAM (percentage of accepted point mutations). In the ALIGN comparison, a bias of 6 was used and gap penalties of 6 and 12 were imposed (1). For

the RELATE and SEARCH analyses, segment lengths of 25 residues were used.

RESULTS AND DISCUSSION

Isolation and Sequence of the ADH Gene. The cDNA (pADH.cDNA) clone used to screen the genomic DNA library of *Drosophila* has been shown to contain sequences coding for the COOH-terminal half of the ADH^F protein (see Fig. 2) (11). By using this DNA as a probe, a clone from the λ library of *Drosophila* genomic DNA (14) with a 15-kilobase (kb) insert containing the ADH gene was identified. *EcoRI* digestion of the λ DNA gave a 4.8-kb fragment to which the probe hybridized (Fig. 1). The location and orientation of the ADH gene within the 4.8-kb fragment were established by hybridization of the cDNA clone and labeled message to the various restriction digests.

By using the strategy depicted in Fig. 1, the nucleotide sequence of the ADH^S gene and flanking regions were determined. Almost 2 kb of DNA was sequenced. In the present paper only the coding region (Fig. 2) of the gene and the introns will be discussed.

Reliability of the Sequence. The sequence given in Fig. 2 is based on three independent sets of data: DNA sequence of the pADH.cDNA clone; DNA sequence of a portion of the genomic 4.8-kb *EcoRI* fragment; and amino acid sequence and composition of several purified tryptic peptides. Moreover, the codon change (ACG to AAG) responsible for the amino acid replacement threonine to lysine observed in the ADH^F and ADH^S allelic variants (20) is found. And lastly, the translated sequence agrees with a recently published amino acid sequence (21) except for our identification of a glutamic acid at position

FIG. 2. DNA sequence for the Adh^S gene of *D. melanogaster*. The initiation and termination codons are indicated by boxes. The translated protein sequence is indicated below the nucleotide sequence. →, Residues observed in tryptic peptides analyzed by automatic Edman degradation (16, 17); ****, tryptic peptide whose composition agrees with that predicted from the DNA; solid vertical arrow, location of the two introns; open vertical arrow, 5' terminus of the cDNA clone (pADH.cDNA) derived from Adh^F mRNA. The dashed line encloses the second-base codon change responsible for the threonine-to-lysine change observed between ADH^F and ADH^S (20). Six silent third-base substitutions are observed between the ADH^S and the COOH-terminal segment of ADH^F .

Table 1. Codon usage in the *Drosophila* ADH gene

No.	Codon	Amino acid	No.	Codon	Amino acid	No.	Codon	Amino acid	No.	Codon	Amino acid
1	TTT	Phe	0	TCT	Ser	1	TAT	Tyr	0	TGT	Cys
8	TTC	Phe	5	TCC	Ser	5	TAC	Tyr	2	TGC	Cys
0	TTA	Leu	0	TCA	Ser	1	TAA	OC	0	TGA	OP
3	TTG	Leu	2	TCG	Ser	0	TAG	AM	5	TGG	Trp
0	CTT	Leu	0	CCT	Pro	1	CAT	His	0	CGT	Arg
3	CTC	Leu	8	CCC	Pro	3	CAC	His	5	CGC	Arg
0	CTA	Leu	2	CCA	Pro	0	CAA	Gln	0	CGA	Arg
21	CTG	Leu	1	CCG	Pro	7	CAG	Gln	0	CGG	Arg
9	ATT	Ile	3	ACT	Thr	2	AAT	Asn	0	AGT	Ser
14	ATC	Ile	22	ACC	Thr	14	AAC	Asn	2	AGC	Ser
0	ATA	Ile	0	ACA	Thr	2	AAA	Lys	0	AGA	Arg
1	ATG	Met	2	ACG	Thr	16	AAG	Lys	0	AGG	Arg
3	GTT	Val	6	GCT	Ala	5	GAT	Asp	6	GGT	Gly
9	GTC	Val	15	GCC	Ala	7	GAC	Asp	8	GGC	Gly
0	GTA	Val	1	GCA	Ala	1	GAA	Glu	5	GGA	Gly
10	GTG	Val	1	GCG	Ala	9	GAG	Glu	0	GGG	Gly

25 and an additional tryptophan at position 251. We are confident in our assignment at these positions because we have isolated tryptic peptides whose sequences agree exactly with those predicted from the DNA at these positions (see Fig. 2).

Codon Usage. There is a distinct bias in the use of degenerate codons for many genes (22, 23). In mammals, for example, C and G are often used as the third codon base whereas A and U are preferred in viruses (23). The codon frequencies in the translation of the *Drosophila* ADH^s gene (Table 1) show preferential use of codons with C or G as the third base. A similar bias is observed with the major heat-shock protein from *Drosophila* (24). From the small amount of data available, it appears that *Drosophila* may show choices among synonymous codons similar to what is observed with mammals (25).

Location of the Introns. We found two untranslated regions in the protein encoding portion of the ADH^s gene (Fig. 2). The first, 65 bases, interrupts the protein sequence at amino acid residue 32, whereas the longer (70 bases) second intron interrupts the protein sequence at amino acid residue 167. Two introns have also been observed by Goldberg (26) using S1 nuclease mapping experiments. The location of intron 2 is substantiated by its absence from the nucleotide sequence of the pADH.cDNA clone. Moreover, both introns contain the 5' G-T and 3' A-G dinucleotides (27, 28) and the consensus sequence (29) (Fig. 3) found at the coding-intron junction of many eukaryotic genes.

Because the x-ray crystallographic structure is not known for the *Drosophila* ADH, the significance of the introns' locations within the gene is not immediately evident. The existence and location of functional domains must be based either on amino acid sequence homology with other dehydrogenases whose pri-

mary and tertiary structures are known or on a search for a recognizable pattern of secondary structural elements predicted from the amino acid sequence.

Sequence Comparison. To assess homology among the alcohol dehydrogenases, we compared the yeast, horse liver, and *Drosophila* ADH protein sequences. The yeast and horse liver sequences were highly related. The significance of such comparisons can be assessed from an alignment score. The alignment score is the number of standard deviations by which the maximal score for the sequences compared exceeds the average maximal score for random permutations of these sequences. An alignment score of 3 is usually taken as an indication of probable relatedness. For the yeast and horse liver sequences, a score of 15.1 was found. The alignment obtained by this analysis was essentially that proposed by Jörnval (30); we found 78 of the 83 amino acid identities he observed. When the *Drosophila* sequence was aligned with either the yeast or the horse liver sequence, the alignment scores were 1.13 and -1.01, respectively.

A similar analysis using the computer program RELATE, which makes an exhaustive comparison of all segments of a given length (25 residues in this case) from one sequence with those from the other, again found little similarity between *Drosophila* ADH and the other two proteins. Moreover, when the *Drosophila* sequence was compared to itself, no duplicated sequences were found. Finally, to test whether the *Drosophila* enzyme might be related to some other protein, 25-residue segments of the *Drosophila* sequence were compared with all the 25-residue segments of proteins or peptides currently in the data bank. At the time the analysis was done, more than 1400 sequences were available. In only one case was a significant

FIG. 3. The DNA at the exon-intron boundaries of the *Drosophila* ADH gene. Vertical broken lines show where the excision-ligation events must occur to give mature mRNA. The underlined nucleotides are the consensus sequence found at the coding-intron junction of many eukaryotic genes (29).

similarity found, involving residues 231–255 of the *Drosophila* ADH and residues 344–368 of rabbit phosphorylase. The reason for this is unclear.

Secondary Structure Prediction. Although there is little sequence homology among the dehydrogenases, they all possess a coenzyme-binding domain of extremely similar structure (12). The domain consists of two roughly identical units with the main elements being six strands of parallel sheet (βA , βB , βC , βD , βE , and βF) and four helices (αB , αC , αE , and $\alpha 1F$). There are two helices on each side of the sheet.

To search for such a structure in the protein sequence of *Drosophila* ADH, the secondary structure of the enzyme was predicted by the Chou–Fasman procedure (18). Overall, nine α -helical regions, 10 β -sheet regions, and 15 β -turn structures are predicted (Fig. 4). The percentage helix present in this structure (47.3%) is in good agreement with that calculated from the CD spectrum of the protein (51.7%).

Starting with residue 8, the predicted structure of *Drosophila* ADH resembles the alternating β -sheet/ α helix arrangement found in the coenzyme-binding domain of dehydrogenases (12). This is seen more clearly in Fig. 5 where the predicted β sheet structures of the *Drosophila* ADH are aligned with the β sheets in the coenzyme binding domain of horse liver ADH. Perhaps more importantly, the four amino acid residues found in all coenzyme binding domains (12, 13) are also present in *Drosophila* ADH. In all dehydrogenases, the last residue in βA is invariably a glycine; any larger residue would interfere with the adenine ribose position (12). An invariant aspartate forms a hydrogen bond to the 2'-OH of the same ribose. A second invariant glycine in αB is close to the AMP phosphate. The third invariant glycine assumes the same relative position at the nicotinamide ribose as does the first one at adenosine. Each one of these residues can be found at the same position in *Drosophila* ADH.

Based on these observations, we believe that the first 140 residues of the *Drosophila* ADH are involved in binding of the coenzyme. This differs considerably from the horse liver enzyme, in which the first 170 residues include the Zn-binding residues (31). The *Drosophila* enzyme does not have any bound metals (32).

In support of our assignment of the coenzyme binding domain to the NH_2 -terminal half of the enzyme is our observation that ADH from the ethyl methanesulfonate-generated mutant *Adh*ⁿ¹¹ cannot bind coenzyme (33). Except for a charge difference, ADH from *Adh*ⁿ¹¹ resembles the wild-type enzyme in every physical property examined (33). The enzyme from this mutant has been shown to have a glycine-to-aspartate change at position 14 (34), the position where replacement by any larger residue would interfere with adenine binding.

If we now reexamine the location of the introns in the *Drosophila* ADH gene, we find that the longer of the two is found

near the boundary of the coenzyme-binding domain. It appears to divide the coenzyme-binding domain from what we assume to be the catalytic domain. The smaller intron, however, is in a β -sheet structure in what we predict is the adenine mononucleotide binding site. In fact, it separates the two invariant glycines in this unit from the invariant aspartate. Thus, whereas the location of one intron supports the idea that exons represent functional units in a protein, the location of the other intron raises serious doubts about the hypothesis.

A similar situation has been observed in the chicken ovomucoid gene (35). Three of the introns delineate the three functional trypsin-binding domains of the secreted ovomucoid polypeptide as well as a separate domain that contains the signal peptide. The domains delineated by these introns are similar to those initially suggested from the protein sequence (36). A second set of introns, however, divide these domains. To explain this finding, it is assumed that the primordial gene consisted only of the 5' subdomain and that the intron was created in the course of the addition of the 3' subdomain (35).

We are therefore left with the problem of defining what constitutes a domain in a protein. Such a problem was apparent with the globins until it was noted that the amino acid sequence encoded in the central exon contained all the necessary contact residues that determine the heme-binding site (3). In fact, this region of the globin polypeptide can bind heme tightly and specifically (37). It has been suggested that this "miniglobin" encoded in the central exon might be a remnant of an early heme carrier (3). Perhaps a similar case can be made for the 5' subdomains of ovomucoid. However, if we accept the x-ray crystallographic data which suggest that this protein family of NAD-dependent dehydrogenases is very old (12) and that the primordial dehydrogenase contained a mononucleotide binding domain represented by the alternating $\beta\alpha\beta$ structure observed in extant dehydrogenases (38), then the location of the intron within this domain of the *Drosophila* ADH is not consistent with this exon being a remnant of a nucleotide binding protein.

These first 32 amino acid residues of *Drosophila* ADH do not appear to possess any unique features that would require such a division from the rest of the protein sequence. Although the β -sheet structure of residues 9–13 has the highest nucleation probability of all the β -sheets, it is only speculative that the nucleation site for the coenzyme binding fold resides in this region. It should be noted, however, that the predicted structure for this region bears a striking resemblance to the secondary structure of a synthetic 34-residue polypeptide designed to bind nucleotides that was shown to possess strong interaction with cytidine phosphates and single-stranded DNA (39).

It is possible that the smaller intron is a recent insertion into the *Drosophila* gene and therefore may not represent the ancestral genetic unit. However, from the work available on the globin (40), ovomucoid (35), and insulin genes (41), the acqui-

FIG. 4. Secondary structure prediction for *Drosophila* ADH by the computerized method of Chou and Fasman (18). α -Helix, 47.3%; β -sheet, 24.2%; random coil, 28.5%. Percentage α -helix calculated from the CD spectrum of *Drosophila* ADH is $51.8 \pm 2.6\%$ (mean $\pm$ SD).

FIG. 5. Alignment of the β -sheet regions in the NAD^+ -binding domain of horse LADH with the predicted β -sheets in residues 8–140 of *Drosophila* ADH (*D. m.* ADH). The adenine mononucleotide binding fold (Upper) is aligned with the nicotinamide mononucleotide binding fold (Lower) below. The four invariant residues found in all dehydrogenases (12) are indicated by white letters in black boxes. The helical section (αCD) connecting βC and βD in horse LADH (residues 250–258) is also found to be helical in the *Drosophila* enzyme (residues 68–83).

sition of introns appears to be an early evolutionary event, perhaps concomitant with the formation of a functional ancestral gene.

The final assignment of structural and functional domains for the *Drosophila* ADH may have to await x-ray crystallographic study of the protein. However, the data presented here suggest that the introns in the *Drosophila* ADH gene do not clearly assort the nucleotide sequence into functional elements. We must await the isolation and analysis of other dehydrogenase genes, especially those that code for proteins whose tertiary structures are known, to see if the conclusions presented here can be generalized.

We thank Peter Chou for performing the computerized secondary structure prediction of the ADH protein. This work was supported in part by National Institutes of Health Grants ES-01527 and GM-28791 and Contract EY-76-S-02-2965 from the Department of Energy to W.S.; C.B. was a Postdoctoral Fellow of the Damon Runyon–Walter Winchell Cancer Fund (DRG-175-F).

- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, 5 (National Biomedical Research Foundation, Washington, DC), Suppls. 1–3, 1973, 1976, and 1978, Vol. 5.
- Gilbert, W. (1978) *Nature (London)* **271**, 501.
- Gilbert, W. (1979) in *Eukaryotic Gene Regulation*, ICN–UCLA Symposium on Molecular and Cellular Biology, eds. Axel, R., Maniatis, T. & Fox, C. F. (Academic, New York), Vol. 14, pp. 1–12.
- Darnell, J. E., Jr. (1978) *Science* **202**, 1257–1260.
- Doolittle, W. F. (1978) *Nature (London)* **272**, 581–582.
- Crick, F. (1979) *Science* **204**, 264–271.
- Blake, C. C. F. (1979) *Nature (London)* **277**, 598.
- Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1485–1489.
- Sakano, H., Rogers, J. H., Hueppi, K., Brack, C., Trautnecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277**, 627–633.
- Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 857–861.
- Benyajati, C., Wang, N., Reddy, A., Weinberg, E. & Sofer, W. (1980) *Nucleic Acids Res.*, in press.
- Rossmann, M. G., Liljas, A., Brändén, C. I. & Banaszak, L. J. (1975) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 11, pp. 61–102.
- Eventoff, W. & Rossmann, M. G. (1975) in *CRC Critical Reviews in Biochemistry*, ed. Fasman, G. D. (CRC Press, Cleveland, OH), pp. 111–140.
- Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- Powers, D. A., Fishbein, J. C., Place, A. R. & Sofer, W. (1980) in *Methods in Peptide and Protein Sequence Analysis*, ed. Birr, C. (Elsevier/North-Holland, Amsterdam), pp. 89–102.
- Fishbein, J. C., Place, A. R., Ropson, I. J., Powers, D. A. & Sofer, W. (1980) *Anal. Biochem.*, in press.
- Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.
- Siegel, J. B., Steinmetz, W. E. & Long, G. E. (1980) *Anal. Biochem.* **104**, 160–167.
- Fletcher, T. S., Ayala, F. J., Thatcher, D. R. & Chambers, G. K. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5609–5612.
- Thatcher, D. R. (1980) *Biochem. J.* **187**, 875–886.
- Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. (1980) *Nucleic Acids Res.* **8**, 49–62.
- Grantham, R., Gautier, C. & Gouy, M. (1980) *Nucleic Acids Res.* **8**, 1893–1912.
- Ingolia, T. D., Craig, E. A. & McCarthy, B. J. (1980) *Cell* **21**, 669–679.
- Grantham, R. & Gautier, C. (1980) *Naturwissenschaften* **67**, 93–94.
- Goldberg, D. A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5794–5798.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
- Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y. & Brownlee, G. G. (1978) *Nature (London)* **275**, 510–513.
- Lerner, M., Boyle, J., Mount, S., Wolin, S. & Steitz, J. (1980) *Nature (London)* **283**, 220–224.
- Jörnvall, H. (1977) *Eur. J. Biochem.* **72**, 443–452.
- Brändén, C.-I. (1977) in *Pyridine Nucleotide-Dependent Dehydrogenases*, ed. Sund, H. (de Gruyter, New York), pp. 325–334.
- Place, A. R., Powers, D. A. & Sofer, W. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **39**, 1640.
- Place, A. R., Powers, D. A. & Sofer, W. (1979) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **38**, 497.
- Retzios, A. D. & Thatcher, D. R. (1979) *Biochimie* **61**, 701–704.
- Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R. & O'Malley, B. W. (1980) *Cell* **21**, 681–687.
- Kato, I., Kohr, W. J. & Laskowski, M. J. (1978) in *Proceedings of the 11th FEBS Meeting*, eds. Magnuson, S. M., Ottesen, B., Taltmann, B., Dano, K. & Neurath, H. (Pergamon, New York), Vol. 47, p. 197.
- Craik, C. S., Buchman, S. R. & Beychok, S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1384–1388.
- Rossmann, M. G., Garavito, R. M. & Eventoff, W. (1977) in *Pyridine Nucleotide-Dependent Dehydrogenases*, ed. Sund, H. (de Gruyter, New York), pp. 3–28.
- Gutte, B., Däumigen, M. & Wittschieber, E. (1979) *Nature (London)* **281**, 650–655.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
- Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20**, 555–565.