

Chicken Riboflavin-binding Protein

cDNA SEQUENCE AND HOMOLOGY WITH MILK FOLATE-BINDING PROTEIN*

(Received for publication, March 7, 1988)

Dao Bin Zheng‡, Heon Man Lim§, Jacques J. Pène§, and Harold B. White III‡¶

From the ‡Department of Chemistry and Biochemistry and the §School of Life and Health Sciences, University of Delaware, Newark, Delaware 19716

The *Rd* gene is expressed in the livers and oviducts of laying hens and codes for the riboflavin-binding protein (RfBP) of egg yolk and egg white. A λ gt11 cDNA library derived from chicken oviduct poly(A)⁺ RNA was screened with polyclonal rabbit antiserum to chicken RfBP. Positive clones were isolated and re-screened with a mixed oligonucleotide probe corresponding to residues 20–25 of the mature protein. The largest cDNA clone (969 base pairs) was subcloned into plasmid pIBI21, and the nucleotide sequence was determined by the dideoxynucleotide method. This clone contained the entire coding region for RfBP. The published amino acid sequence of the mature protein was confirmed. In addition, the following 17-residue signal peptide was deduced: Met-Leu-Arg-Phe-Ala-Ile-Thr-Leu-Phe-Ala-Val-Ile-Thr-Ser-Ser-Thr-Cys. Unexpectedly, the nucleotide sequence codes for 2 adjacent arginine residues at the carboxyl terminus that are not observed in the mature protein. The amino acid sequence of RfBP is homologous with bovine milk folate-binding protein. Eight of the nine pairs of cysteines involved in disulfide bonds in RfBP are conserved in folate-binding protein, as are all of the tryptophan residues. Sequence identity between homologous regions of these two vitamin-binding proteins is more than 30%.

The faint yellow color of normal egg white is due to riboflavin bound to a specific riboflavin-binding protein (1). This protein (RfBP),¹ constituting 0.8% of egg white protein (2), is also present in egg yolk (3, 4) and in the plasma of laying hens (5). Hens homozygous for the recessive *rd* allele are unable to produce RfBP and lay eggs deficient in riboflavin (6). Embryos in these eggs die at or about 13 days of development (7).

Mature RfBP is the product of a number of post-translational modifications. In addition to the cleavage of a heretofore uncharacterized signal peptide, its amino terminus is blocked by pyroglutamic acid (8), and an acidic carboxyl-

terminal 11–13-residue peptide is removed during or after transport to the yolk (9). There are nine disulfide bonds whose positions have been determined (10), two complex, *N*-linked oligosaccharides whose composition depends on the tissue of synthesis (5, 8), and a serine-rich region containing eight phosphoryl groups (11–14). These phosphoryl groups are necessary for the transport of serum RfBP into the oocyte (15). The mature 219-amino acid sequence has been determined (8, 9). Crystals have been obtained, and structural analysis is in progress (16).

RfBP is unusual among egg proteins because it is present in both egg yolk and egg white (1, 17). This implies that the estrogen-dependent expression of RfBP is distinct from that of other egg proteins such as egg white ovalbumin, synthesized in the oviduct (18, 19), or vitellogenin, synthesized in the liver (20).

Homologous riboflavin-binding proteins are present in the eggs of reptiles (21) and other birds (22–24) and in the plasma of pregnant mammals (24, 25). In the later case, RfBP has been shown to be necessary for the transfer of riboflavin from mother to fetus. Immature female rats immunized with chicken RfBP subsequently develop autoimmunity to their pregnancy-specific RfBP and spontaneously abort (26).

As a first step toward understanding the transcriptional regulation of the *Rd* gene, determining the molecular basis of the null allele, and providing a probe for isolating the mammalian gene, a cDNA clone coding for chicken RfBP has been isolated, and its nucleotide sequence was determined.

MATERIALS AND METHODS

cDNA Library—The cDNA library was obtained from Dr. Bert W. O'Malley's laboratory at Baylor Medical School (Houston, TX). The construction of this library in λ gt11 from chick oviduct poly(A)⁺ RNA has been described (27).

Screening for RfBP-producing Clones—Propagation of recombinant λ gt11 was conducted in *Escherichia coli* strain γ 1090(r[−]) (28) obtained from Promega Biotec (Madison, WI). Clones were screened with the Photoblot Immunoscreening System by the method of the supplier (46). Rabbit antibodies to chicken RfBP were obtained from Dr. Robert Guyer (Pennsylvania State University, University Park, PA). Clones expressing RfBP antigens were isolated, and the recombinant λ gt11 DNA was purified (29). The sizes of the cDNA inserts were determined by gel electrophoresis of *Eco*RI-digested DNA in 1% Sea Kem GTG-agarose gels (FMC Bioproducts, Rockland, ME). The largest inserts were rescreened for hybridization with a mixed oligonucleotide probe (30), synthesized on an Sysmec Microsyn DNA synthesizer. The probe was chosen to be complementary to a region with low degeneracy near the amino terminus of RfBP (residues 20–25).

DNA Sequence Determination—The clone containing the largest insert (~1 kilobase pair) and also hybridizing with the oligonucleotide probe was selected for sequencing. The cDNA was excised from λ gt11 and subcloned (31) into the *Eco*RI site of pIBI21. DNA from recombinant plasmids was isolated (32) and used as a template in the DNA sequencing reaction (32, 33). The M13 17-mer reverse primer (pur-

* This work was supported by the University of Delaware Research Foundation, a National Institutes of Health Biomedical Research Support Group grant, and Abbott Laboratories. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03922.

¶ To whom correspondence should be addressed.

¹ The abbreviations used are: RfBP, riboflavin-binding protein; FBP, folate-binding protein.

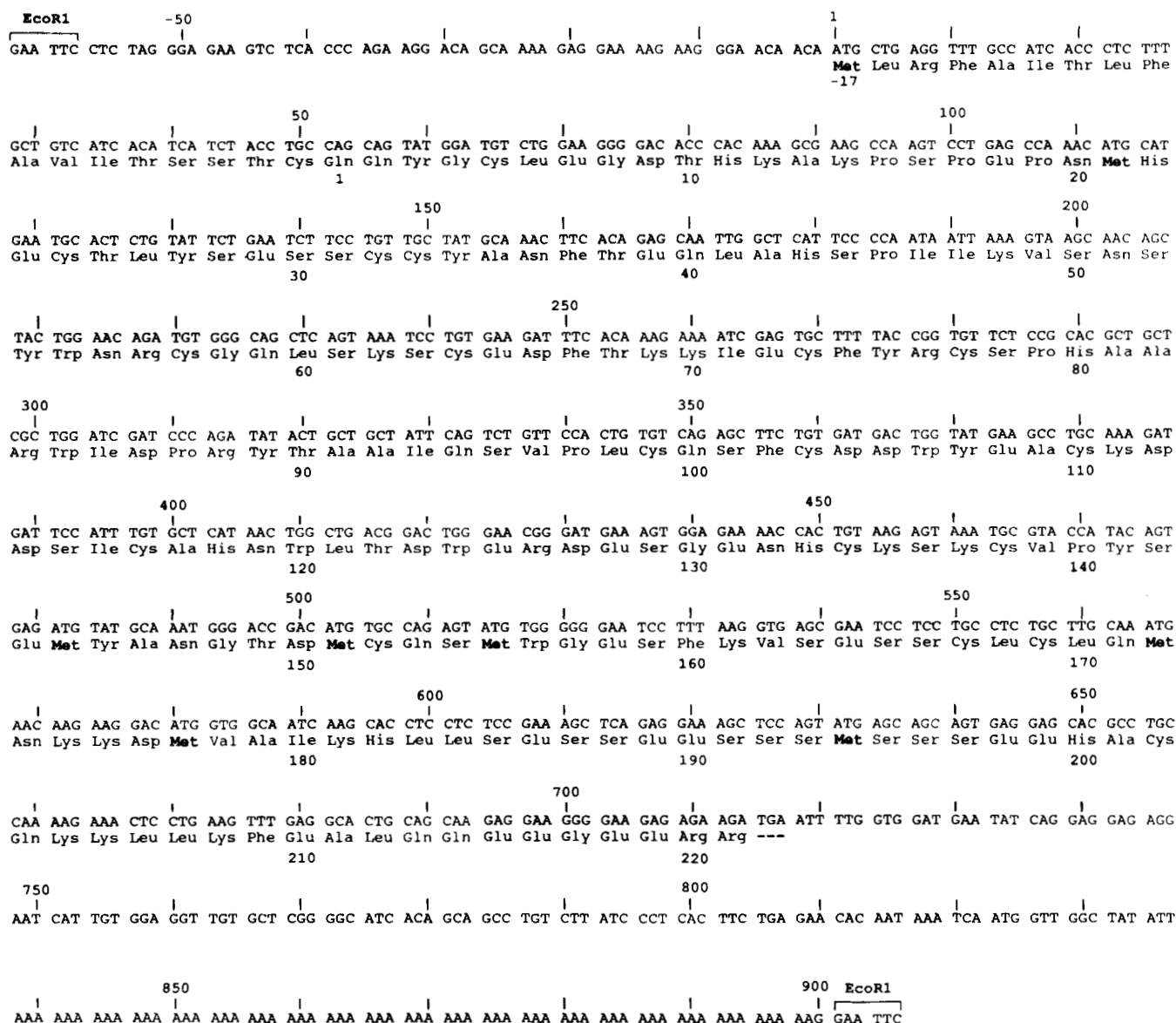


FIG. 1. Nucleotide sequence of cDNA clone coding for chicken riboflavin-binding protein. Numbers above the nucleotide sequence are referenced to the initiation codon, whereas numbers below the amino acid sequence are referenced to the amino terminus of the mature protein.

chased from Pharmacia LKB Biotechnology Inc.) was used to obtain the nucleotide sequence of the 5' end of the gene. Subsequent sequencing reactions were positioned by primers synthesized to correspond to distal regions of the sequence determined in the preceding sequence reaction. The nucleotide sequences of both strands were determined using a total of nine different primers.

RESULTS AND DISCUSSION

Propagation and Detection of Rd cDNA Clones—Riboflavin-binding protein is one of a number of proteins in egg white that have or may have antimicrobial activity (34). Consequently, production of significant amounts of RfBP in *E. coli* could be lethal or bacteriostatic because the protein could scavenge exogenous and *de novo* synthesized riboflavin. To avoid this potential selection against the clones of interest, the culture media for propagation and screening were supplemented with riboflavin. This precaution is analogous to that used in screening avidin-producing clones from the same cDNA library by Gope *et al.* (27).

A total of 30 clones were independently isolated using a polyclonal rabbit antiserum to chicken RfBP. All but four of

these λ gt11 clones contained cDNA inserts of fewer than 700 base pairs and thus were not large enough to encode the 219 amino acids in RfBP and its expected signal peptide. Only the largest of the four remaining clones hybridized with a mixed nucleotide probe corresponding to residues 20–25 of the mature protein. The nucleotide sequence of this clone included the entire coding region of the *Rd* gene.

This full-length *Rd* cDNA subcloned in an expression vector did not impair the growth of host cells on unsupplemented media, suggesting that the expressed fusion protein was produced in very small amounts or was unable to bind riboflavin. This conclusion is consistent with the fact that RfBP is a secreted protein with many disulfide bonds. Furthermore, its expression in chicken involves a number of other post-translational modifications unexpected in an intracellular fusion protein in *E. coli*. The other termination codon (TAG) shortly following the 5'-*EcoRI* site (positions -54 to -52) suggests that no fusion protein would be produced. However, screening with antiserum indicates that fusion protein was produced perhaps in small quantities as a result of the amber-ocher

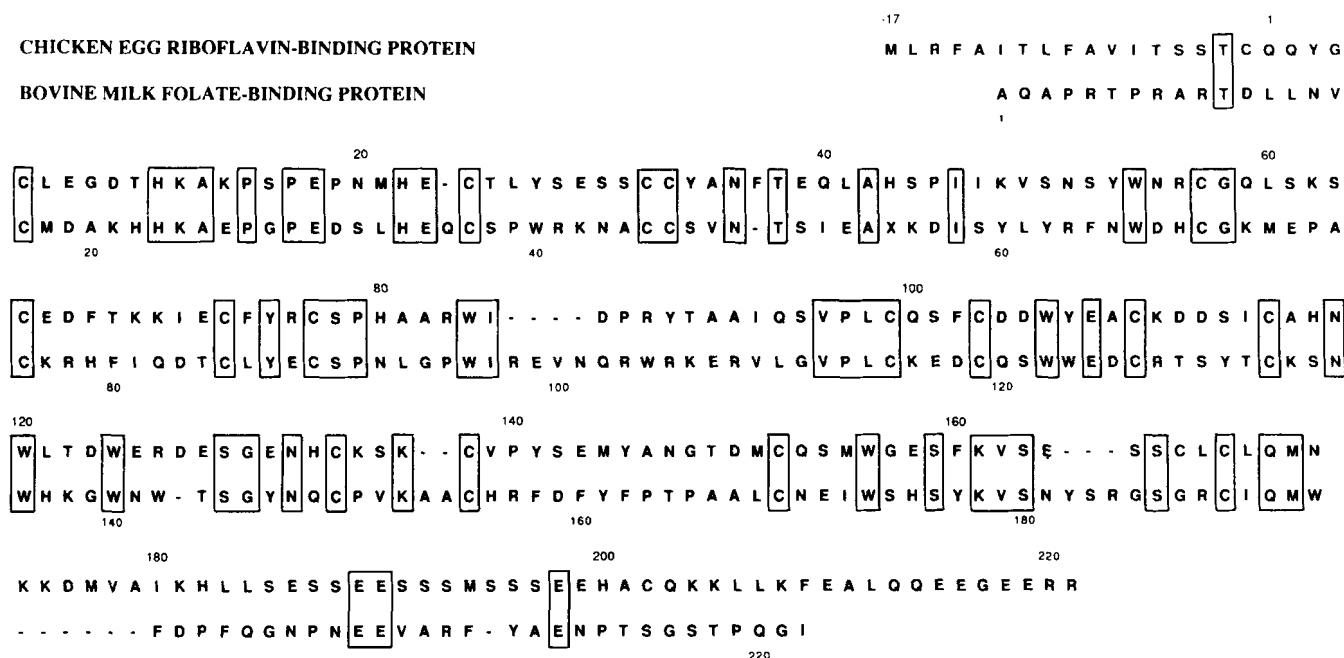


FIG. 2. Alignment of amino acid sequence of chicken riboflavin-binding protein (8, 9) and bovine folate-binding protein (39). Numbers above and below the sequences correspond to residue positions in the mature proteins. Sequence identities are boxed. Residue 55 of the folate-binding protein is undetermined. The order of residues 155 and 156 in FBP is uncertain (39). The order selected here preserves the positions of cysteine residues in both sequences.

suppressor in the *E. coli* γ 1090 host strain.

Nucleotide Sequence Determination—The nucleotide sequence of the *Rd* cDNA was determined by the dideoxynucleotide method using oligonucleotide primers that successively extended the sequence. Both strands of the cDNA were sequenced. In the region coding for the signal peptide, the nucleotide sequence was difficult to determine. This problem was overcome, and the nucleotide sequence was confirmed by using a primer that abutted this region.

The nucleotide sequence (Fig. 1) is in complete agreement with the amino acid sequence determined by Hamazume *et al.* (8). The polymorphism they observed in the protein sequence at residue 14 indicated that the predominant allele coded for lysine at that position. This is the allele we have isolated. A 17-amino acid signal peptide is predicted by the nucleotide sequence. The predicted site for signal peptide cleavage precedes adjacent glutamine residues. The resulting amino-terminal glutamine must cyclize, eliminating ammonia, to produce the pyroglutamate found in the mature protein.

Although the published amino acid sequence of RfBP (8, 9) agrees with the nucleotide sequence reported here, the nucleotide sequence indicates that the initial translation product contains adjacent arginine residues at the carboxyl terminus. These are not observed in the mature protein; and thus, another peptide processing event must occur in addition to those mentioned in the Introduction. A similar observation has been made with the cDNA sequence for the matrix γ -carboxyglutamic acid containing protein of bone, dentin, and cartilage, where a carboxyl-terminal Arg-Arg-Gly-Ala-Lys sequence is predicted from the nucleotide sequence, but is not observed in the mature protein (35).

Recent analysis of the nucleotide sequences flanking the ATG initiation codon of almost 700 vertebrate cDNAs reveals the consensus sequence GCCGCCPuCCATGG for initiation of translation (36). Other than the ATG initiation site, the

Rd cDNA matches this sequence at positions -2, -3, -5, and -9. This 5'-untranslated region of the *Rd* cDNA is distinctive in having a polypurine tract just upstream of the initiation codon (positions -25 to -6). Another polypurine tract occurs immediately preceding the termination codon (positions 692-714). Higher than expected frequencies of polypurine tracts of this size and greater have been found in the human β -globin region and several other human gene sequences (37). The significance of these regions is unclear.

The 3'-untranslated region of *Rd* cDNA is 120 nucleotides long, with the AATAAA consensus sequence for polyadenylation at positions 814-819 (38). The poly(A) tract in this clone contains 62 adenyl residues.

Homology with Milk Folate-binding Protein—A search of a protein sequence data bank by Dr. Russell Doolittle (University of California, San Diego, CA) revealed a significant similarity of RfBP to a fragment of bovine milk folate-binding protein (FBP) (39). Subsequent comparison with the completed FBP sequence (40) has extended the region of apparent homology. Alignment to maximize identical residues in the two sequences indicates that RfBP has a carboxyl-terminal extension relative to FBP, whereas FBP is extended at the amino terminus (Fig. 2). In the region from residues 5 to 172, where the sequence clearly corresponds to that of FBP, there are several small insertions or deletions. Sequence identity in this region is more than 30%, a value that greatly exceeds the null hypothesis that the identities are due to chance. The presumption of homology is supported by structural and functional considerations.

RfBP has nine disulfide bonds. The specific cysteines involved in each disulfide have been determined (10). All but one of these pairs of cysteine residues are conserved in FBP, suggesting that the disulfide bonds in the two proteins are probably the same. Tryptophan residues have been implicated in riboflavin binding to RfBP (41). All 6 of the tryptophan residues in RfBP are conserved in FBP. One of the two *N*-

linked oligosaccharides in each protein occurs at a conserved asparagine residue. These conserved residues account for 23 of the 59 sequence identities between the two vitamin-binding proteins.

Notably absent in the FBP sequence is the highly anionic region at the carboxyl terminus of RfBP. In RfBP, this region includes a 14-residue sequence that contains 8 phosphorylserine and 5 glutamic acid residues (11, 14). Whereas this region is not found in milk FBP, similar sequences are well known in caseins (42), the major proteins of milk. Whether RfBP acquired this phosphorylated region or FBP lost it in their respective evolution from a common ancestral protein may become evident when the structure of chicken egg FBP is determined. This recently discovered vitamin-binding protein² is one of a number of nutrient-transport proteins now known from chicken egg (43).

The recognition of homology between RfBP and FBP establishes another family of high affinity, soluble, transport proteins. Previously, it has been shown that vitamin D-binding protein is homologous with albumen and α -fetoprotein (44) and that retinol-binding is homologous with bilin-binding protein and milk β -lactoglobulin (45). Whereas folates and flavins are functionally distinct in their catalytic function in cells, they do have structural resemblance in the pterin ring system they share. Thus, it is conceivable that an ancestral protein had the capacity to bind both vitamins or that a few amino acid replacements could interconvert the binding specificities.

Acknowledgments—We thank Dr. Bert W. O'Malley for the gift of the hen oviduct cDNA library, Dr. Robert Guyer for antiserum to chicken RfBP, and Dr. Russell Doolittle for searching the protein data bank for homologies to RfBP and for aligning the RfBP and FBP sequences.

REFERENCES

- White, H. B., III, and Merrill, A. H., Jr. (1988) *Annu. Rev. Nutr.* **8**, 279–299
- Rhodes, M. B., and Bennett, N. and Feeney, R. E. (1959) *J. Biol. Chem.* **234**, 2054–2060
- Zak, Z., and Ostrowski, W. (1963) *Acta Biochim. Pol.* **10**, 427–441
- White, H. B., III, Armstrong, J., and Whitehead, C. C. (1986) *Biochem. J.* **238**, 671–675
- Miller, M. S., Bruch, R. C., and White, H. B., III (1982) *Biochim. Biophys. Acta* **715**, 126–136
- Winter, W. P., Buss, E. G., Claggett, C. O., and Boucher, R. V. (1967) *Comp. Biochem. Physiol.* **22**, 897–906
- Maw, A. J. G. (1954) *Poult. Sci.* **33**, 216–217
- Hamazume, Y., Mega, T., and Ikenaka, T. (1984) *J. Biochem. (Tokyo)* **95**, 1633–1644
- Norioka, N., Okada, T., Hamazume, Y., Mega, T., and Ikenaka, T. (1985) *J. Biochem. (Tokyo)* **97**, 19–28
- Hamazume, Y., Mega, T., and Ikenaka, T. (1987) *J. Biochem. (Tokyo)* **101**, 217–223
- Fenselau, C., Heller, D. N., Miller, M. S., and White, H. B., III (1985) *Anal. Biochem.* **150**, 309–314
- Vaughn, V. L., Wang, R., Fenselau, C. and White, H. B., III (1987) *Biochem. Biophys. Res. Commun.* **147**, 115–119
- Miller, M. S., Mas, M. T., and White, H. B., III (1984) *Biochemistry* **23**, 569–576
- Mega, T., Hamazume, Y., Nong, Y. M., and Ikenaka, T. (1986) *J. Biochem. (Tokyo)* **100**, 1109–1116
- Miller, M. S., Benore-Parsons, M., and White, H. B., III (1982) *J. Biol. Chem.* **257**, 6818–6824
- Zanette, D., Monaco, H. L., Zanotti, G., and Spadon, P. (1984) *J. Mol. Biol.* **180**, 1185–1187
- Durga Kumari, B., and Adiga, P. R. (1986) *J. Biosci.* **10**, 193–202
- O'Malley, B. W., and Means, A. R. (1974) *Science* **183**, 610–620
- Grody, W. W., Schrader, W. T., and O'Malley, B. W. (1982) *Endocr. Rev.* **3**, 141–163
- Tata, J. R., James, T. C., Watson, C. S., Williams, J. L., and Woffle, A. P. (1983) *Molecular Biology of Egg Maturation* (Porter, R., and Whelan, J., eds) pp. 96–110, Pitman Medical Publishing Co., Ltd., Kent, Great Britain
- Abrams, W. A. M., McGahan, T. J., Rohrer, J. S., Bero, A. S., and White, H. B., III (1988) *Comp. Biochem. Physiol.* **90B**, 243–247
- Feeney, R. E., and Allison, R. G. (1969) *Evolutionary Biochemistry of Proteins*, Wiley-Interscience, New York
- Muniyappa, K., and Adiga, P. R. (1980) *Biochem. Biophys. Res. Commun.* **623**, 339–347
- Adiga, P. R., and Ramana Murty, C. V. (1983) *Molecular Biology of Egg Maturation* (Porter, R., and Whelan, J., eds) pp. 111–131, Pitman Medical Publishing Co., Ltd., Kent, Great Britain
- Merrill, A. H., Jr., Froehlich, J. A., and McCormick, D. B. (1979) *J. Biol. Chem.* **254**, 9362–9364
- Ramana Murty, C. V., and Adiga, P. R. (1982) *Science* **216**, 191–193
- Gope, M. L., Keinonen, R. A., Dristo, P. A., Conneely, O. M., Beattie, W. G., Zarucki-Schulz, T., O'Malley, B. W., and Kulomaa, M. S. (1987) *Nucleic Acids Res.* **15**, 3595–3606
- Young, R. A., and Davis, R. W. (1983) *Science* **222**, 778–782
- Huynh, T., Young, R. A., and Davis, R. W. (1983) in *DNA Cloning Techniques: A Practical Approach* (Glover, D., ed.) p. 49, IRL Press, Oxford
- Suggs, S. V., Hirose, T., Miyake, T., Kawashime, E. H., Johnson, M. J., Itakura, K., and Wallace, R. B. (1981) *ICN-UCLA Symp. Mol. Cell Biol.* **23**, 682–693
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, p. 391, Academic Press, New York
- Lim, H. M., and Pène, J. J. (1988) *Gene Anal. Techn.* **5**, 32–39
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Tranter, H. A., and Board, R. G. (1982) *J. Appl. Biochem.* **4**, 295–338
- Price, P. A., Fraser, J. D., and Metz-Virca, G. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8335–8339
- Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148
- Behe, M. J. (1987) *Biochemistry* **26**, 7870–7875
- Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263**, 211–214
- Svendsen, I., Martin, B., Pedersen, T. G., Hansen, S. I., Holm, J., and Lyngbye, J. (1979) *Carlsberg Res. Commun.* **44**, 89–99
- Svendsen, I., Hansen, S. I., Holm, J., and Lyngbye, J. (1984) *Carlsberg Res. Commun.* **49**, 123–131
- Blankenhorn, G. (1978) *Eur. J. Biochem.* **82**, 155–160
- West, D. W. (1986) *J. Dairy Res.* **53**, 333–352
- White, H. B., III (1987) *J. Exp. Zool.* **1**, (suppl.) 53–63
- Cooke, N. E. (1986) *J. Biol. Chem.* **261**, 3441–3450
- Sawyer, L. (1987) *Nature* **327**, 659
- Promega Biotec (1986) *Promega Technical Bulletin* 32, Madison, WI

² T. Sherwood and H. B. White III, unpublished data.