

Rabbit Muscle Creatine Phosphokinase

cDNA CLONING, PRIMARY STRUCTURE, AND DETECTION OF HUMAN HOMOLOGUES*

(Received for publication, August 2, 1984)

Scott Putney‡§, Walter Herlihy¶, Nancy Royal¶, Henrianna Pang¶, H. Vasken Aposhian‡, Lulu Pickering‡, Rama Belagaje||, Klaus Biemann¶, David Page‡, Stephen Kuby**, and Paul Schimmel‡

From the Departments of ‡Biology and ¶Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, ||Eli Lilly, Inc., Indianapolis, Indiana 46240, and the **Department of Biochemistry and Laboratory for Study of Hereditary and Metabolic Diseases, University of Utah Research Park, Salt Lake City, Utah 84108

A cDNA library was constructed from rabbit muscle poly(A) RNA. Limited amino acid sequence information was obtained on rabbit muscle creatine phosphokinase and this was the basis for design and synthesis of two oligonucleotide probes complementary to a creatine kinase cDNA sequence which encodes a pentapeptide. Colony hybridizations with the probes and subsequent steps led to isolation of two clones, whose cDNA segments partially overlap and which together encode the entire protein. The primary structure was established from the sequence of two cDNA clones and from independently determined sequences of scattered portions of the polypeptide. The reactive cysteine has been located to position 282 within the 380 amino acid polypeptide. The rabbit cDNA hybridizes to digests of human chromosomal DNA. This reveals a restriction fragment length polymorphism associated with the human homologue(s) which hybridizes to the rabbit cDNA.

Creatine phosphokinase is prominent in energy metabolism of muscle tissue where it generates ATP from creatine phosphate and ADP. There are two isozymes: M, which is found as the MM dimer in skeletal muscle and as the MB heterodimer in cardiac tissue; and B, which is present as a BB dimer in brain and in many other tissues (1). The M-isozyme has been extensively investigated from the standpoint of mechanism and structure-function (see Refs. 1-9 and references therein). In spite of many investigations, detailed interpretations are prevented by the lack of a primary structure.

For the purpose of determination of the primary structure and to gain access to creatine kinase genes, we attempted to

* This work was supported by National Institutes of Health Grants GM05472 and RR00317 (K. B.) and grants from the American Heart Association (P. S.) and Muscular Dystrophy Association (P. S.). 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.

§ Postdoctoral fellow supported by the American Cancer Society.

isolate a muscle creatine kinase cDNA clone. cDNA clones for 13 muscle-specific proteins were previously identified in a rabbit muscle cDNA library by shotgun sequencing (10). The translated sequences of randomly selected clones were compared with known amino acid sequences of muscle-specific proteins. The more abundant muscle-specific cDNA clones were identified by their respective sequences. The paucity of known creatine phosphokinase amino acid sequences decreased sharply the effectiveness of the approach for detecting the M-isozyme cDNA. To identify clones of rabbit muscle creatine phosphokinase from a cDNA library, synthetic oligonucleotides complementary to the creatine phosphokinase cDNA were used as specific hybridization probes to screen the library.

The cDNA library was constructed in the *Pst*I site of pBR322 by addition of dC tails to mRNA:cDNA hybrids (synthesized by reverse transcriptase extension from an oligo(dT) primer (11)), annealing to dG-tailed pBR322 and transforming *Escherichia coli* strain C600. Approximately 1600 ampicillin-sensitive tetracycline-resistant clones were obtained from 5.0 μ g of rabbit muscle poly(A) mRNA.

Two deoxyoligonucleotides, each 14 nucleotides long, were used as hybridization probe to screen and identify M creatine phosphokinase clones in a cDNA library constructed from rabbit muscle poly(A) RNA. As the basis for selecting the deoxyoligonucleotide probe sequences, experiments were designed to obtain a peptide sequence with two closely spaced methionine residues. Rabbit M creatine phosphokinase was cleaved with cyanogen bromide and the sequence of the resultant fragments was determined by GC/MS.¹ (12, 13). Because peptides containing more than approximately six amino acids are not detected by GC/MS and because cyanogen bromide cleaves after methionine (14), any peptide sequence obtained is likely to have two methionine residues closely spaced in the amino acid sequence. These experiments, and an independent determination (by Edman degradation) of the sequence of a fragment isolated from a proteolytic digestion, established the presence of the peptide Leu-Met-Val-Glu-Met-Glu.

Possible codons encoding Met-Val-Glu-Met-Glu were considered, as well as codon usage patterns of rabbit muscle mRNA (10). Two deoxyoligonucleotides were made by the phosphotriester method (15, 16): probe I, 5' TCCATCTCCACCAT; and probe II, 5' TCCATTTCAAC-CAT.

Both oligonucleotides were labeled at the 5'-end and hybridized by the method of Wallace *et al.* (17) to a cDNA library constructed in the *Pst*I site of pBR322. Of 1600 independent colonies, 14 contain inserts that hybridize to both probes. Further analysis led to isolation of two clones which collectively contain the entire creatine phosphokinase coding sequence. These clones are designated pCKM15 and pCKM19. Both plasmids have a 180-bp *Pst*I fragment; pCKM15 has sequences 3' to this fragment (including the poly(A) tract), while pCKM19 has the cDNA 5' to the frag-

¹ The abbreviations used are: GC/MS, gas chromatography/mass spectrometry; FAB/MS, fast-atom bombardment/mass spectrometry; SSC, sodium chloride and sodium citrate buffer (3 M NaCl, 41 mM sodium citrate, pH 7.0); kb, kilobase; bp, base pair.

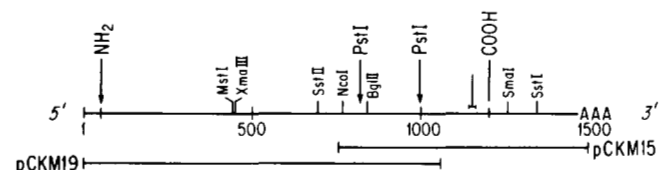


FIG. 1. Map of M creatine phosphokinase cDNA showing location of clones pCKM15 and pCKM19. The coding region begins 55 nucleotides from the 5' side and ends 268 nucleotides from the poly(A) tract at the 3' side. These points are indicated with

arrows showing the amino and carboxyl termini. The location hybridizing to the oligonucleotide probes (encoding Met-Val-Glu-Met-Glu) is shown by the overbar, and the location of the 180-bp *Pst*I fragment (which encodes the reactive cysteine) is also indicated. The regions of the cDNA contained by pCKM15 and pCKM19 are indicated by the brackets.

The cDNA and amino acid sequence shown in Fig. 2 for

arrows showing the amino and carboxyl termini. The location hybridizing to the oligonucleotide probes (encoding Met-Val-Glu-Met-Glu) is shown by the overbar, and the location of the 180-bp *Pst*I fragment (which encodes the reactive cysteine) is also indicated. The regions of the cDNA contained by pCKM15 and pCKM19 are indicated by the brackets.

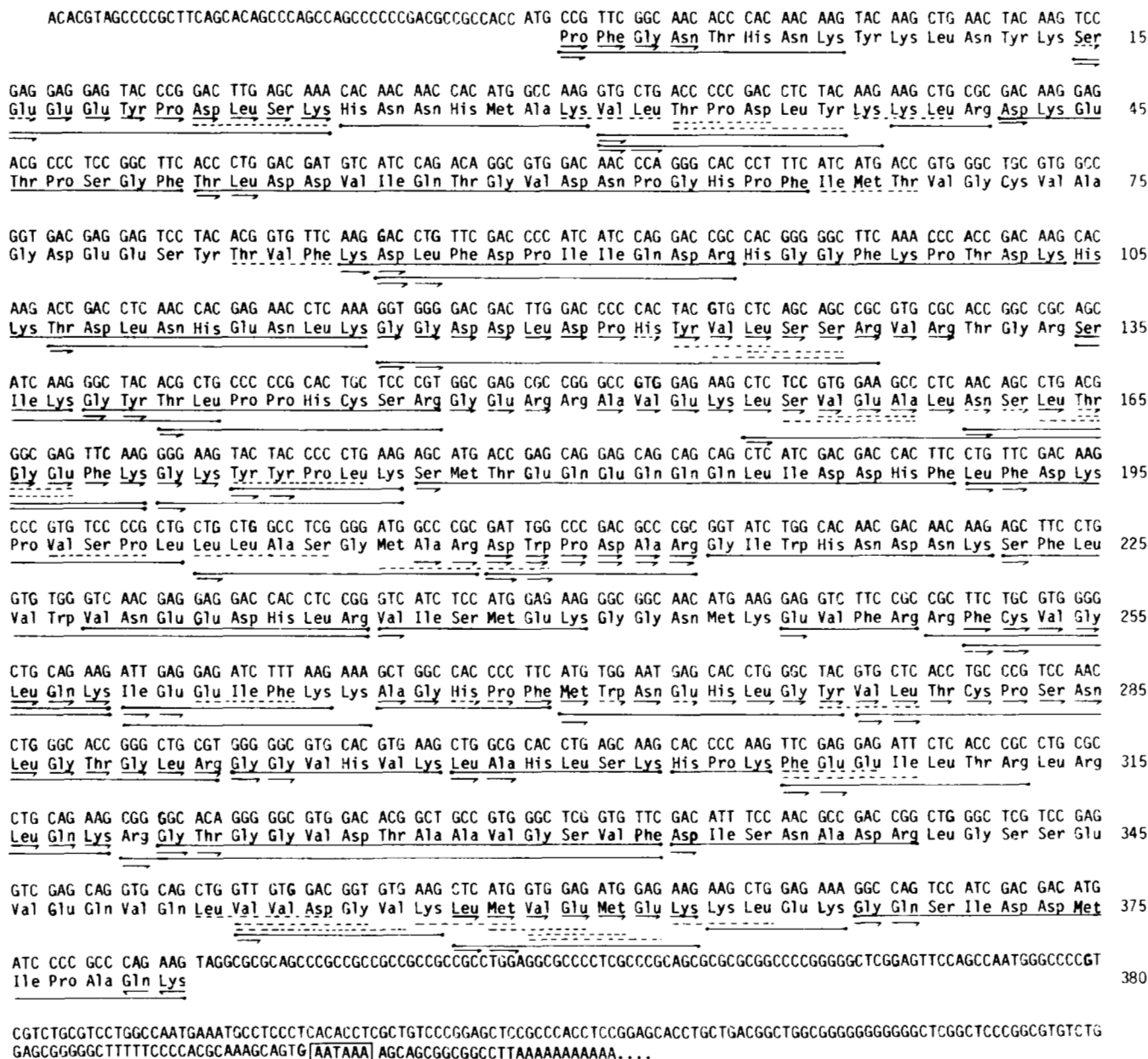


FIG. 2. Rabbit muscle creatine phosphokinase cDNA sequence and translated polypeptide sequence. The coding region begins with proline at amino acid +1. The poly(A) addition signal sequence AATAAA is boxed. Dashed lines are placed under peptides identified by GC/MS sequencing. Successive Edman degradations on internal fragments and the N terminus are designated by arrows. (Dashed arrows indicate uncertainty in an Edman experiment). Solid underlines are placed under peptides identified by FAB/MS. The Edman shortened peptides analyzed by FAB/MS are shown by arrows immediately underneath the solid lines. Rabbit muscle creatine phosphokinase was cleaved by cyanogen bromide, elastase, and thermolysin, or proteinase K for GC/MS analysis (12, 13). The protein was carboxymethylated, cleaved by trypsin (with some chymotryptic activity) for FAB/MS analysis (22) using the procedure developed by Gibson and Biemann (21). The molecular weights of the tryptic and chymotryptic fragments were determined, in addition, the exact mass values (with an accuracy of 4 millimass units) of 21 of these fragments were obtained to confirm their identities. The fragments were also subjected to Edman degradation(s) and the molecular weights of the shortened peptides were determined by FAB/MS (21).

rabbit M creatine phosphokinase is the first complete sequence of this enzyme to be determined from any organism. Several criteria confirm that the translated cDNA sequence corresponds exactly to that of the protein. 1) The first ATG from the 5'-end (at nucleotide 52) begins an open reading frame that ends with the termination codon TAG at nucleotide 1195, while each of the other two reading frames contains a number of termination codons. In the open reading frame, the polypeptide, beginning with a proline encoded by CCG after the initiator ATG, contains 380 amino acids with a calculated molecular weight of 42,932. This compares well with that determined by other physical methods (20). 2) 19% of the polypeptide sequence is confirmed by short amino acid sequences determined by GC/MS and distributed throughout the polypeptide sequence. Tryptic and chymotryptic peptides covering 86% of the sequence were matched (21) by molecular weight as determined by FAB/MS (22). There were no peptides detected by FAB/MS which did not match any of the predicted ones. Edman degradation of some tryptic peptides confirmed 26% of the sequence. Altogether 90% of the translated sequence was confirmed independently. 3) The amino terminus of the polypeptide is confirmed by the detection of the predicted N-terminal tryptic peptide by FAB/MS and four Edman steps on the whole protein. Carboxypeptidase digestion revealed Gln-Lys as the C terminus of the creatine phosphokinase (23) as predicted by the DNA sequence. In addition, the C-terminal tryptic dodecapeptide was detected by FAB/MS. 4) Three independent determinations of the amino acid composition of rabbit muscle creatine phosphokinase (20, 24, 25) agree within 15% or better with that calculated for the protein structure deduced from the DNA sequence.

Creatine phosphokinase contains a reactive cysteine which is at or near the active site (1, 4). Cysteines are located in the primary structure at positions 73, 145, 253, and 282. By comparing the partial sequence of the labeled peptides (26, 27) containing the reactive cysteine with the sequence in Fig. 2, we located the reactive cysteine to position 282. Recent experiments suggest that three histidines are at or near the active site (8). Note that His-268, His-275, and His-295 are close to Cys-282. Having established the sequence, the precise identification of these histidines and of the active-site carboxylate (6, 28) and arginine (29) is now possible.

The canonical sequence AATAAA (30) occurs 15 nucleotides prior to the poly(A) tract (Fig. 2). RNA blot hybridization, using the 180-bp *Pst*I fragment as a probe, shows the length of the message encoding M creatine phosphokinase to be about 1600 nucleotides.² The length of the cDNA is 1470 base pairs and, assuming the poly(A) tract is 100 to 150 nucleotides in length, this would make the cDNA the same size as the message.

DNA from eleven unrelated humans was digested with *Eco*RI, *Hind*III, *Msp*I, or *Taq*I and probed with either the 0.18-kb *Pst*I fragment or the 0.8-kb probe from pCK19 (Fig. 1). With either probe, between three and eight fragments were observed with each of the enzymes. In general, fragments observed with the 0.18-kb probe were a subset of those visualized with the 0.8-kb probe. These results suggest that there are at least three loci in the human genome homologous to rabbit M creatine phosphokinase.

Fig. 3 shows hybridization of the 0.8-kb fragment to *Taq*I digested DNA from four subjects. Eight bands are observed, two of which hybridize more intensely than the others. (Bands of differing intensities were also seen on blots with the other

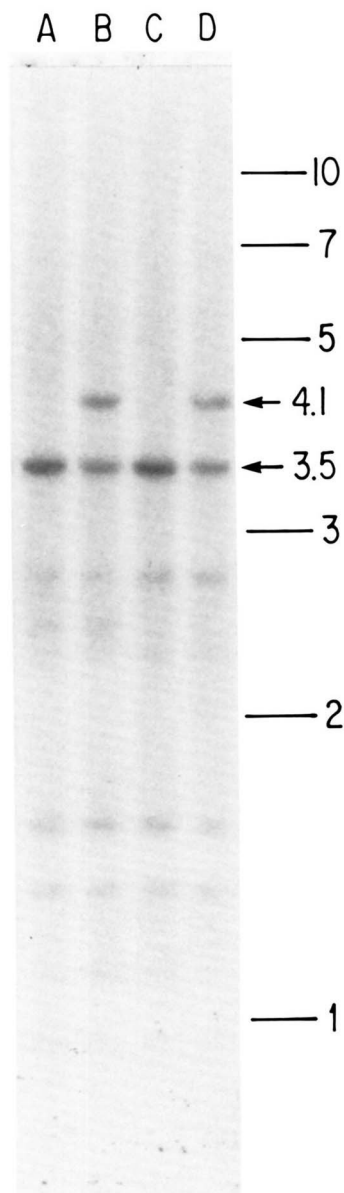


FIG. 3. Hybridization of rabbit M creatine phosphokinase to human DNA. Human DNA from four persons (lanes A-D) were digested with *Taq*I, electrophoresed, blotted as previously described (31), and probed with the 0.8-kb *Pst*I fragment from pCKM19. Eight bands are visible, two of which (3.5 and 4.1 kb) are more intense and show restriction fragment length polymorphism. Hybridization was done in $5 \times$ SSC, 50% formamide at 42 °C, and washes were in $1 \times$ SSC, 0.1% sodium dodecyl sulfate at 37 °C.

enzymes.) A restriction fragment length polymorphism (32) is also evident. Individuals in lanes B and D display a 4.1-kb fragment not seen in lanes A and C. A corresponding decrease in the intensity of the 3.5-kb *Taq*I band in lanes B and D suggests that the 3.5- and 4.1-kb fragments are allelic. Independent polymorphisms were observed in *Eco*RI and *Msp*I digests (not shown).

Altogether, at least three independent restriction fragment length polymorphisms were observed, with digests of *Eco*RI, *Msp*I, and *Taq*I. These polymorphisms may be of use as genetic markers (32). Note that hybridization of the rabbit cDNA to human DNA occurs under moderately stringent conditions. Because M creatine phosphokinase from various species is similar (as shown by partial sequences³ and by

² L. Pickering and P. Schimmel, unpublished observations.

³ S. Kuby, unpublished observations.

immunological cross-reactivity (33, 34)), it is not surprising that specific human fragments are detected with the rabbit probe. In all experiments, bands of two different hybridizing intensities were seen. The intense bands may be due to hybridization to the M-isozyme creatine kinase gene while the weaker bands may contain the B-isozyme genes. Hybridization of the rabbit M and B creatine phosphokinase cDNAs has been observed.²

The rabbit muscle creatine kinase sequence reported here is the first complete cDNA and amino acid sequence of creatine kinase from any source. Clones of chicken muscle creatine kinase cDNAs have been reported (35, 36) and a partial cDNA and amino acid sequence of rat muscle creatine kinase has been published (37). The latter lacks 5'-noncoding and amino-terminal coding sequences. The 295 codons which have been determined are exactly colinear with the rabbit sequence reported here, with homology of the partial nucleotide and amino acid sequences of greater than 90%. The 3'-noncoding sequences diverge, and the rabbit sequence has 73 additional nucleotides in the 3'-noncoding region.

Acknowledgment—We thank Dr. Joseph Avurich of Massachusetts General Hospital for the N terminus analysis.

REFERENCES

- Watts, D. C. (1973) in *The Enzymes* (Boyer, P. D., ed) Vol. 8, pp. 383-455, Academic Press, New York
- Hammes, G. G., and Huest, J. K. (1969) *Biochemistry* **8**, 1083-1094
- Schimerlik, M. I., and Cleland, W. W. (1973) *J. Biol. Chem.* **248**, 8418-8423
- Smith, D. J., and Kenyon, G. L. (1974) *J. Biol. Chem.* **249**, 3317-3318
- McLaughlin, A. C., Leigh, J. S., Jr., and Cohn, M. (1978) *J. Biol. Chem.* **251**, 2777-2787
- Marletta, M. A., and Kenyon, G. L. (1979) *J. Biol. Chem.* **254**, 1879-1886
- Vasak, M., Nagayama, K., Wuthrich, K., Mertens, M. L., and Kagi, J. H. R. (1979) *Biochemistry* **18**, 5050-5055
- Rosevear, P. R., Desmeules, P., Kenyon, G. L., and Mildvan, A. S. (1981) *Biochemistry* **20**, 6155-6164
- Hsu, H., and Watanabe, J. (1983) *Chem. Pharm. Bull (Tokyo)* **31**, 3269-3276
- Putney, S. D., Herlihy, W. C., and Schimmel, P. (1983) *Nature (Lond.)* **302**, 718-721
- Roninson, I. B., and Ingram, V. M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4782-4785
- Herlihy, W. C., Royal, N. J., Biemann, K., Putney, S. D., and Schimmel, P. R. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 6531-6535
- Carr, S. A., Herlihy, W. C., and Blemann, K. (1981) *Biomed. Mass Spectrom* **8**, 51-61
- Gross, E., and Witkop, B. (1962) *J. Biol. Chem.* **237**, 1856-1860
- Narang, S. A., Brousseau, R., Hsiung, H. M., and Michniewicz, J. J. (1980) *Methods Enzymol.* **65**, 610-620
- Narang, S. A., Hsiung, H. M., and Brousseau, R. (1979) *Methods Enzymol.* **68**, 90-98
- Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., and Itakura, K. (1981) *Nucleic Acids Res.* **9**, 879-894
- Messing, J., and Vieira, J. (1982) *Gene* **19**, 269-276
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467
- Yve, R. H., Palmieri, R. H., Olson, O. E., and Kuby, S. A. (1967) *Biochemistry* **6**, 3204-3227
- Gibson, B. W., and Biemann, K. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 1956-1960
- Barber, M., Bordoli, R. S., Sedgwick, R. D., and Tyler, A. N. (1981) *Nature (Lond.)* **293**, 270-275
- Olson, O. E., and Kuby, S. A. (1964) *J. Biol. Chem.* **239**, 460-467
- Noltman, E. A., Mahowald, T. A., and Kuby, S. A. (1962) *J. Biol. Chem.* **237**, 1146-1154
- Eppenberger, H. M., Dawson, D. M., and Kaplan, N. D. (1967) *J. Biol. Chem.* **242**, 204-209
- Thomson, A. R., Eveleigh, J. W., and Miles, J. B. (1964) *Nature (Lond.)* **203**, 267-269
- Mahowald, T. A. (1965) *Biochemistry* **4**, 732-740
- Cook, P. F., Kenyon, G. L., and Cleland, W. W. (1981) *Biochemistry* **20**, 1204-1210
- Borders, C. L., Jr., and Riordan, J. F. (1975) *Biochemistry* **14**, 4699-4704
- Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature (Lond.)* **263**, 211-214
- Page, D., deMartville, B., Barker, D., Wyman, A., White, R. L., Francke, U., and Botstein, D. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 5352-5356
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314-331
- Viala, B., Robin, Y., and Van Thoai, N. (1970) *Comp. Biochem. Physiol.* **32**, 401-410
- Bulcke, J. A., and Sherwin, A. L. (1969) *Immunochemistry* **6**, 681-687
- Schweinfest, C. W., Kwiatkowski, R. W., and Dottin, R. P. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4997-5000
- Rosenberg, U. B., Kunz, G., Frischauf, A., Lehrach, H., Mahr, R., Eppenberger, H. M., and Perriard, J.-C. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 6589-6592
- Benfield, P. A., Zivin, R. A., Shearman, C. W., Graf, D., Henderson, L., Oroszlan, S., and Pearson, M. L. (1984) *Exp. Biol. Med.* **9**, 187-194