Complete cDNA sequence for rabbit muscle glycogen phosphorylase

Kenichi Nakano, Peter K. Hwang and Robert J. Fletterick*

Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA

Received 6 June 1986

The cDNA for the nearly full-length rabbit muscle glycogen phosphorylase mRNA has been isolated and sequenced. The cDNA is rich in G and C nucleotides. This feature is especially striking at the 3rd position of codons, where 86% of the 843 amino acid codons terminate with G or C. Methionine, presumably the initiation residue, is found at position −1, suggesting that the removal of only a single methionine residue precedes the amino-terminal acetylation at serine. Eight differences between the deduced amino acid sequence and the previously determined protein sequence are discussed.

Amino acid sequence *Codon usage bias*

1. INTRODUCTION

Glycogen phosphorylase (EC 2.4.1.1), which catalyzes the release of glucose 1-phosphate from stored glycogen, plays a central role in the regulation of intracellular carbohydrate metabolism [1]. Rabbit muscle phosphorylase has been well-characterized and exhibits a complex range of regulatory responses. X-ray crystallographic analysis [2] and the amino acid sequence [3] of this enzyme have led to a three-dimensional model of the protein structure and proposals for catalytic and regulatory mechanisms. A powerful method for testing such proposals and thus advancing our understanding of the molecular mechanisms of phosphorylase function is to analyze the structure and enzymatic properties of appropriately mutated enzyme. With the advent of site-specific metagenesis and gene expression technologies, it has now become feasible to change any amino acid residue of phosphorylase to any other amino acid and to determine the effect of the change. As a prerequisite to such studies, we here report the isolation and sequencing of a cDNA spanning the entire coding region of rabbit muscle phosphorylase.

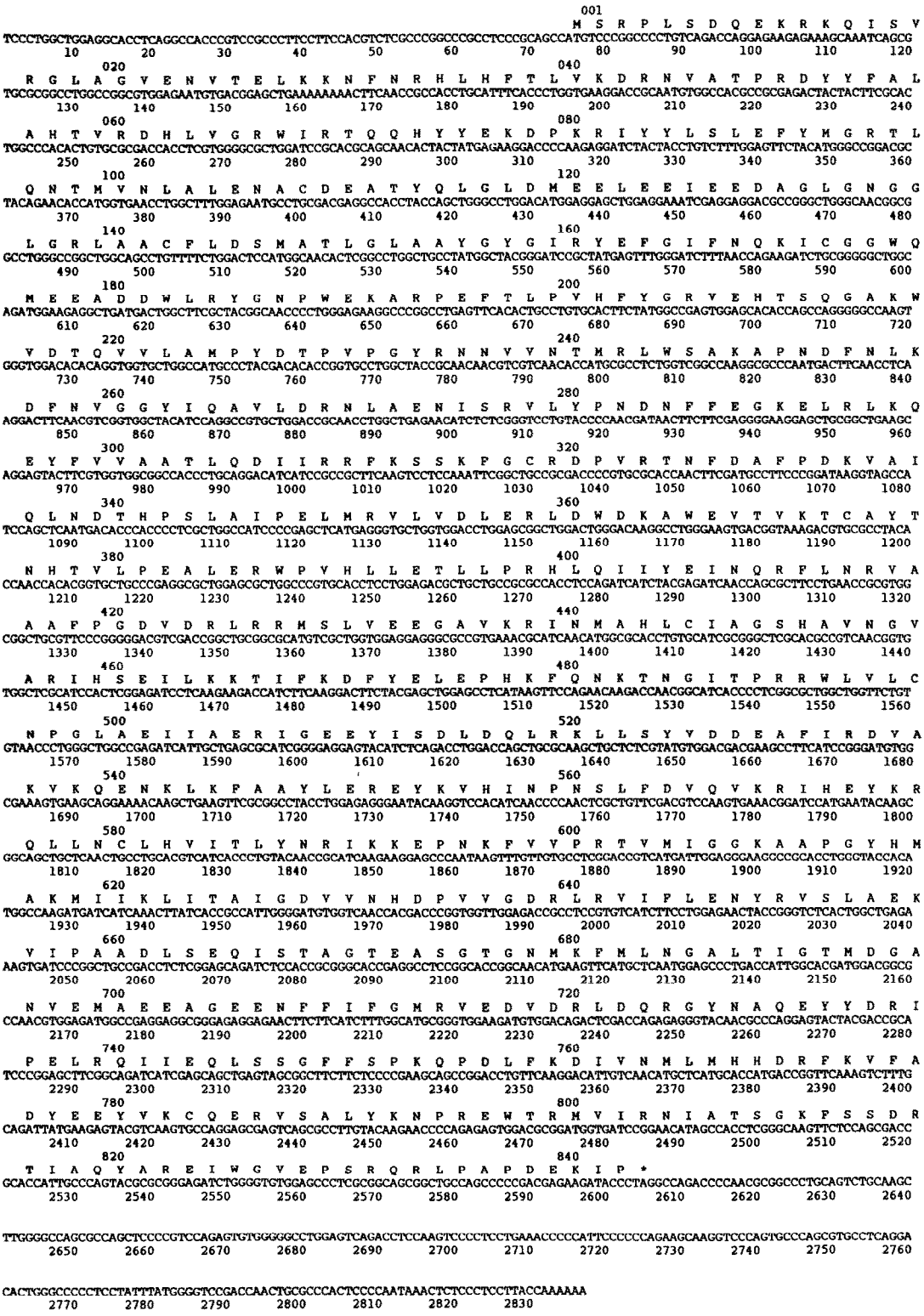
2. MATERIALS AND METHODS

A cDNA library prepared from muscle of 1–3-day-old rabbits, constructed in the vector pcD-X by the method of Okayama and Berg [4], was provided by Dr D.H. MacLennan, University of Toronto, Canada [5]. Screening of this cDNA library was performed according to the colony-screening procedures of Grunstein and Hogness [6]. A cDNA fragment encoding residues 573–743 of rabbit muscle phosphorylase was used as a probe [7] to screen about 20000 transformant colonies. The probe was labeled by nick-translation [8] and hybridized under highly stringent conditions (50% formamide, 42°C, overnight). Plasmids were isolated from colonies hybridizing to the probe, digested separately with *Bam*HI, *Pst*I and *Xho*I restriction endonucleases, and analyzed by agarose gel electrophoresis to estimate the molecular size of the inserted DNA.

Various restriction fragments of the cDNA-containing plasmid were ligated into M13 vectors and DNA sequencing was carried out by the

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50



dideoxy method [9]. To determine the sequence of ambiguous regions, electrophoresis was performed on a 5% polyacrylamide gel containing 25% formamide (deionized) and 7 M urea. After electrophoresis, the gel was dried directly on paper and autoradiographed at room temperature for 2–20 h.

3. RESULTS AND DISCUSSION

A C-terminal cDNA fragment of rabbit muscle phosphorylase was previously isolated [7]. In this study, we screened for full-length phosphorylase clones in another cDNA library using as probe an upstream portion of the C-terminal cDNA, a 0.51 kb fragment encoding residues 574–744. We were able to isolate a plasmid containing about 2.9 kb of insert DNA, which is large enough to contain the entire coding reigon.

Fig.1 shows the restriction map and the sequencing strategy for the insert. During the sequencing, we frequently encountered problems of irregular ladder patterns on gels, presumably due to the formation of secondary structures that do not denature during the gel run. In most cases, sequencing the opposite strand did not improve the results. These difficulties were overcome by adding formamide to the sequencing gels. The resulting formamide gels gave much clearer resolution in the irregular regions of the gel without affecting the other regions, as far as we tested. The amino-terminal 67 amino acids and the 3′-untranslated region were sequenced in only one direction (the reverse sequence), due to the difficulty of sequencing through the poly G(C)-tailed regions at the cloning site at the 5′-side of the vector. To verify the sequence of the 5′-untranslated region, the oligonucleotide covering from methionine (−1) to aspartic acid 6 was synthesized and used as a primer for dideoxy sequencing. All other regions were sequenced in both directions.

The nucleotide sequence of rabbit muscle phosphorylase cDNA is shown in fig.2. This sequence corrects several positions in the cDNA sequence published in [7]. The earlier errors presumably derived from the sequencing difficulties alluded to above. The coding region exhibits a high G+C content (60%), which was a major cause of sequencing difficulties. The anomalous G+C richness is also found in the 5′- and 3′-untranslated regions (76 and 65%, respectively). It is expected that this high G+C content would be a common feature among mammalian genes for skeletal muscle phosphorylase; indeed, partial cDNA sequences of rabbit, rat, and human muscle phosphorylases were found to be highly homologous [7].

The G+C richness in the rabbit muscle phosphorylase cDNA is primarily concentrated at the 3rd position of codons, where 86% are G or C, as shown in table 1. This feature is also observed in cDNA sequences for other rabbit muscle enzymes [10], including creatine kinase [11] and aldolase A [12]. The 5′- and 3′-untranslated regions for these proteins are also enriched in G or C. The biased base compositions in the third position of codons may imply some special functional role for the DNA structure in muscle-specific genes.

The deduced phosphorylase sequence is composed of 843 amino acids. The methionine residue at position −1 probably serves as the amino-terminus of the initially translated polypeptide, since the nucleotide sequence immediately upstream of this position is homologous with the consensus sequence of eucaryotic initiation sites [13]. Hence, the removal of only a single methionine residue would precede the amino-terminal acetylation at serine.

We have compared the phosphorylase amino acid sequence deduced from cDNA sequencing with that determined by protein sequencing [3]. An additional amino acid, isoleucine, has been introduced at position 308 in the amino acid sequence deduced from cDNA. Addition of the isoleucine gives a better homology with amino acid sequences of potato and *Escherichia coli* phosphorylases [14,15]. This insertion is also anticipated from examination of the electron density map of rabbit muscle phosphorylase at 2.1 Å

Fig.1. Restriction map of rabbit muscle glycogen phosphorylase cDNA. The letters mark the sites for cleavage by the restriction enzymes *Bam*HI, *Pst*I, *Kpn*I, *Xma*I, *Sac*I and *Sph*I. Arrows show the positions and sequencing directions for the fragments that were subcloned into M13 vectors and sequenced.

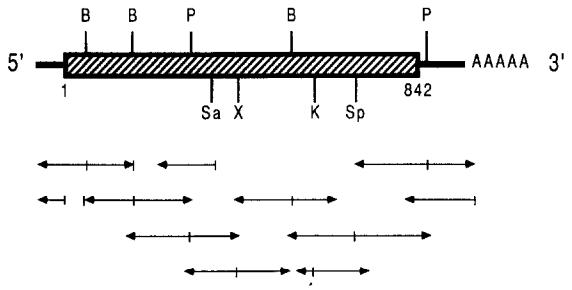


Fig.2. The cDNA and translated protein sequence for rabbit muscle glycogen phosphorylase. The numbering scheme is the same as for the protein sequence for residues 1 (Ser) to 307 (Ile). After Ile 307, the insertion of an isoleucine residue causes the subsequent amino acid sequence shown here to be incremented by one over the previously published protein sequence [3].

resolution (Sprang, S., personal communication).

In addition to the above difference, there are 7 other changes in the amino-terminal portion of the sequence, i.e. positions 30, 32, 42, 55, 57, 88 and 112. Most of these differences are between acidic residues and the corresponding amides, and are likely to be accounted for by errors in protein sequencing. Such discrepancies are occasionally found between primary structures determined from protein and DNA sequencing. In the other cases, histidine and leucine are switched at positions 55 and 57, and aspartic acid at position 112 is found to be a threonine in the DNA sequence. Some of the 8 amino acid changes might be significant in interpreting the function of the enzyme, since their positions are close to the regulatory sites (Asp 42 binds AMP) and form part of the subunit contact region.

X-ray crystallographic analysis at 2.1 Å resolution can be used to verify the amino acid changes found at positions 55, 57 and 112. In each case, the electron density map is in agreement with the amino acid sequence as deduced from the cDNA sequence. Recent analysis on the N-terminal fragment (encoding from the 5′-untranslated region to residue 67) from human muscle phosphorylase gene has also verified our deduced amino acid sequence (Burke, J., Hwang, P.K. and Fletterick, R.J., unpublished). It is noteworthy that the amino acid sequences of human and rabbit muscle phosphorylases are identical from position −1 to 67. As further verification, the corrected residues

Table 1

Codon usage of rabbit muscle glycogen phosphorylase

F	TTT	6	S	TCT	2	Y	TAT	6	C	TGT	2
F	TTC	32	S	TCC	8	Y	TAC	30	C	TGC	7
L	TTA	0	S	TCA	3		TAA	0		TGA	0
L	TTG	3	S	TCG	10		TAG	1	W	TGG	12
L	CTT	3	P	CCT	8	H	CAT	4	R	CGT	1
L	CTC	16	P	CCC	17	H	CAC	18	R	CGC	29
L	CTA	1	P	CCA	1	Q	CAA	3	R	CGA	3
L	CTG	56	P	CCG	10	Q	CAG	27	R	CGG	23
I	ATT	6	T	ACT	1	N	AAT	7	S	AGT	1
I	ATC	42	T	ACC	21	N	AAC	39	S	AGC	5
I	ATA	2	T	ACA	4	K	AAA	9	R	AGA	4
M	ATG	22	T	ACG	10	K	AAG	39	R	AGG	3
V	GTT	3	A	GCT	9	D	GAT	8	G	GGT	3
V	GTC	17	A	GCC	37	D	GAC	41	G	GGC	27
V	GTA	2	A	GCA	5	E	GAA	9	G	GGA	4
V	GTG	40	A	GCG	12	E	GAG	56	G	CGG	14

Note the preference for codons ending in G or C

at positions 32, 42, 55, 57 and 88 are also found to be conserved in the sequence of human liver phosphorylase (Newgard, C., Nakano, K., Hwang, P.K. and Fletterick, R.J., in preparation).

The size of the cDNA reported here (2.9 kb) is still less than the mRNA size (~3.5 kb) estimated by Northern analysis [7]. The discrepancy may be accounted for partially by the poly(A) tail and by a short missing portion of the 5′-untranslated region. Nevertheless, a full cDNA coding region for rabbit muscle glycogen phosphorylase is now fully characterized and available for developing an in vitro expression and mutagenesis system. This work is in progress.

ACKNOWLEDGEMENTS

The authors would like to thank Prudence Bothen for preparation of this manuscript and acknowledge NIH Grant R01 AM32822 for support.

REFERENCES

[1] Graves, D.J. and Wang, J.H. (1972) The Enzymes, 3rd edn, 7, pp.435–482.

[2] Fletterick, R.J. and Madsen, N.B. (1980) Annu. Rev. Biochem. 49, 31–61.
[3] Titani, K., Koide, A., Hermann, J., Ericsson, L.H., Kumar, S., Wade, R.D., Walsh, K.A., Neurath, H. and Fischer, E.H. (1977) Proc. Natl. Acad. Sci. USA 74, 4762–4766.
[4] Okayama, H. and Berg, P. (1983) Mol. Cell. Biol. 3, 280–289.
[5] MacLennan, D.H., Brandl, C.J., Korczak, B. and Green, N.M. (1984) Nature 312, 696–700.
[6] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[8] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning – A Laboratory Manual, Cold Spring Harbor Laboratory, NY.
[9] Sanger, F., Nicklen, S. and Couson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5663–5667.
[10] Putney, S.D., Herliky, W.C. and Schimmel, P. (1983) Nature 302, 718–721.
[11] Putney, S., Herliky, W., Royal, N., Pang, H., Vasken Aposhian, H., Pickering, L., Belagaje, R., Biemann, K., Page, D., Kuby, S. and Schimmel, P. (1986) J. Biol. Chem. 259, 16317–16320.
[12] Tolan, D.R., Amsden, B., Putney, S.D., Urdea, M.S. and Penhoet, E.E. (1984) J. Biol. Chem. 259, 1127–1131.
[13] Kozak, M. (1986) Nucleic Acids Res. 12, 857–872.
[14] Nakano, K. and Fukui, T. (1986) J. Biol. Chem., in press.
[15] Palm, P., Goerl, R. and Burger, K.J. (1985) Nature 313, 500–502.