

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

001
M S R P L S D Q E K R K Q I S V

TCCTGGCTGGAGGCACTCAGGCCACCCGTCGCGCCCTCCCTCCACGCTCTGCGCCGGCCCGCCTCCCGCAGCCATGTCCCGGCCCTGTCCAGACCAGGAGAAAGCAAATCAGCG
10 20 30 40 50 60 70 80 90 100 110 120

020 040

R G L A G V E N V T E L K K N F N R H L H F T L V K D R N V A T P R D Y Y F A L
TGC CGG C C T T G C C G C G T G A A T T G A C G G A C T T G A A A A A A A C T T C A C C G C C A C T T G C A C C C T T G A G G A C C G C A A T G T G C C A C G C C G G A G A C T A C T A C T T C G C A C
130 140 150 160 170 180 190 200 210 220 230 240

060 080

A H T Y R D H L V G R W I R T Q Q H Y E K D P K R I Y Y L S L E F Y H G R T L
T G C C C A C A C T G T C G C G A C C C T C G T G G G C G C T G G A T C C G C A C G A C C A C T A T A T G A A G G A C C C C A A G A G G A T C T A C T A C C T G T T T T G G A G T T C T A C A T G G C G C G A C C
250 260 270 280 290 300 310 320 330 340 350 360

100 120

Q N T M V N L A L E N A C D E A T Y Q L G L D M E E L E E I E E D A G L G N G G
T A C G A A C A C A C G T G A A C T G G C T T T G G A G A A T G C C T G C G A C G A G C C A C C T A C C A C G T G G C C T G G A C A T G A G G A G C T G G A G G A A T C G A E G A G C C C G G C T G G C A A C G G C
370 380 390 400 410 420 430 440 450 460 470 480

140 160

L G R L A C F L D S M A T L G L A A Y G Y G I R Y E F G I F N Q K I C G G W Q
G C T T C A A G C T G C G C G C T T T T C T G C C T C A T G G A C A C A C C T G G C T G A G A A C A T C T C G G G T C G T A C C C C A A C G A T A A C T T T C G A G G G A A G K A G C T G C G G G C T G G C
490 500 510 520 530 540 550 560 570 580 590 600

180 200

M E A A D W L R Y G N P W E K A R P E F T L P Y H F Y G R V E H T S Q G A K W
A G A T T G A A G A G C T G A T G A C T G G C T T C G C T A C G G A A C C C T G G A G A A G C C C G C C T G A G T T C A C T G C C C T G C A C T T C T A T G G C C G A G T G G A G A C C A G C C A G G G C C A A G T
610 620 630 640 650 660 670 680 690 700 710 720

220 240

V D T Q V Y L A M P Y D T P Y P G Y R N N V Y N T M R L W S A K A P N D F N L K
G G T G G C A C A C A G T G T G C C A T G P C A C A C A C C G T G C C T A C C G C A A C A C A C G T G C A A C C A C C A T G G C C T T G G T G C C A A G G C C C A A T G A C T T C A A C C T C A
730 740 750 760 770 780 790 800 810 820 830 840

260 280

D F N V G G Y I Q A V L D R N L A E N I S R V L P N D N F P E G K E L R L K Q
A G G A C T T C A A G C T G C G C G C T A C A T C A G C C C T G G A C C G C A C C T G G C T G A G A A C A T C T C G G G T C G T A C C C C A A C G A T A A C T T T C G A G G G A A G K A G C T G C G G C T G A A G C
850 860 870 880 890 900 910 920 930 940 950 960

300 320

E Y F V Y A A T L Q D I I R R F K S S K P G C R D P V R T N F D A F P D K V A I
A G G A T A C T T C G T G G C G C C C A C C C T G C A G G A C A T A C C C G C C T T C A A G T T C G G C T G C C G C A C C C C G T G C G C A C C A A C T T C G A T G C C T T C C C G G A T A A G T A G C C A
970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080

340 360

Q L N D T H P S L A I C P E L M R Y L V D L E R L D W D K A W E V T Y K T C A Y T
T C C A C T T C A A G C A C C C C C T G C T G C C A T C C C C A G C T A T G A A G G T G C T G G A C C T G G A C C C T G G A C C C G A C C C C G T G G G A C A A G C C T G G G A T A A G C T G C G C C T A C A
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200

380 400

N H T V L P E A L E R W P V H L L E T L P R H L Q I I Y E I N Q R F L N R V A
C C A C C A C C G T G C T G C C C G A G G C G C T G G A C C C T G C C C G C A C C C T G C C A G A T C A T C A G A G A C C A A G C C C T G G G A T A A G C T G C C A A C C G C T G G
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320

420 440

A A P G D V D R L R R H S L V E E G A Y K R I N M A H L C I A G S H A V N G V
C G C T G C T T C C C G G G A C C T G A C C G C C T G C G C C G C A C C C T G C T G T G G A G A G G G C C C G T G A A C C G A T C A A C A T G C G C G C A C T T G C A T G C G G C T C C G A C C C C T C A A C C G T G
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440

460 480

A R I H S E I L K K T I P K D F Y E L E P H K F Q N K T N G I T P R R W L V L C
T G C C A T T C A C T G G A G C T C A A G A C A C C A C T T C A A G G A C T T C A C G A G C C T C A A G T T C C A G A A A A G A C C A A C C G C C A C C C C T G C G C C T G C C T G G T T C G T
1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560

500 520

N P G L A E I A E R I G E E Y I S D L D Q L R K L L S Y T D D E A F I R D Y A
G T A A C C T G G C C G A G A T A T T G C T G A G C C A T C C G G G A G A T A C A T C C A G C C A G C T G C C G A A G C T G C T C T C G T A T G T G Y D D E A F I R D Y A
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680

540 560

K Y K Q E N K L K F A Y L E R E Y K V H I N P N S L F D V Q V K R I H E Y K R
C G A A A G T G A A G C A G A A A C A A G T G A A G T T C G C G C C T A C T G G A G A G G A A T A C A A G G T C C A A C C C C A A C T G C G T T T C G A C V T C C A A G T G A A C C G A T C C A T G A A T C A A C C
1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

580 600

Q L L N C T L H V I T L Y N R I K K E P N K F V Y P R T V M I G G K A A P G Y H M
G G C A C T G C T A A C T G C C T G C A T C A C C C T G T A C A C C G C A T C A A G A G G A C C C C A A T A A G T T G T G C C T C C G A C C C T A T A G T T G G A G G A A G C C C G C A C C T G C C T A C C A C A
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

620 640

A K M I I K L I T A I G D V Y N H D P V Y G D R L R V I P L E N Y R V S L A E K
T G C C A A G T A T A C A A C T T A C C C G C A T T G G G A T G T G T C A A C C A C G A C C C G T G T G G A G A C C G C C T C C G T G C A T C T C C T G G A G A C C A C C G G G T C C A C C G G T T C A C T G G C T G A G A
1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040

660 680

V I P A A D L S E Q I S T A G T E A S G T G N M K F M L N G A L T I G T M D G A
A A G T G A T C C C G C T G C C G A C C T C T C G A G C A G A T C C A C C G C G C A C C G A G C C T G G C A C A C T G A A G T T C A T G C T C A A T G G A G C C A C C A C C C T G G C A C C A T T G G C A C G A T G G A C C G G C
2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160

700 720

N V E M A E E A G E E N F F I F G M R V E D V D R L D Q R G Y C Q E Y Y D R I
C C A A C G T G G A G A T G G C C G A G G G C G G A G A G G A A C T T C T C A T C T T T G G C A T G C G G T G G A A G A T G T G G A C A G A C T C G A C C A G A G G G T A C A A C C C A G G A T A C T A C G C C G C A
2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280

740 760

P E L R Q I E Q L S S G F F S P K Q P D L F K D I V N M L M H H D R F K V F A
T C C C G C A T T C C G C A G A T C A T C G A G C A G T G A T G C G G C T T C T C T C C C G A A G C C C G G A C C T T C A A G G A C A T G T C A A C T M C L M H H D R F K V F A
2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

780 800

D Y E E Y V K C Q E R V S A L Y K N P R E W T R M V I R N I A T S G K F S S D R
C A G A T T A T A G A G T A C T A A G T G C C A G G A C G A T C A G C C L T T G T A C A G A A C C C C A G A G A T G G C G G A T G G T G A T C C G G A A C T A G C C A C C T G C G G A G T T C T C C A G C C G A C C
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

820 840

T I A Q Y A R E I W G V E P S R Q R L P A P D E K I P *
G C A C A T T G C C C A G T A C C G C G G G A T T G G G T T G G A C C C T C G C G G C A C C G G C T C C A C C C C G A C G A G A T A C C C T A G C C A G C C C C A A C C G C G C C T G C A G T T G C A A G C
2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

TGGGGCCAGCCGACCTCCCGTCCAGAGTGTGGGGCCCTGGAGTACAGACTTCAAGTCCCTCCCTGAAACCCCAATTCCTCCCGCAGAAGCAAGTCCCAATGCCAGCGTCCCTCAGGA
2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760

CACTGGGCCCCCTCTATTTATGGGGTCCGACCAACTGGCGCCACTCCCAATAAACTCCCTCCCTTACCAAAAAA
2770 2780 2790 2800 2810 2820 2830

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 region.

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.

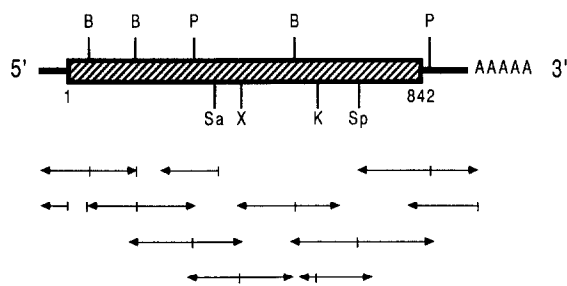


Fig.2. The cDNA and translated protein sequence for rabbit muscle glycogen phosphorylase. The numbering scheme is the same as for the protein sequence for residues 1 (Ser) to 307 (Ile). After Ile 307, the insertion of an isoleucine residue causes the subsequent amino acid sequence shown here to be incremented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