BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **220,** 72–77 (1996)
ARTICLE NO. 0359

The Human Skeletal Muscle Glycogenin Gene: cDNA, Tissue Expression, and Chromosomal Localization

F. Barbetti, M. Rocchi,* M. Bossolasco, R. Cordera,† P. Sbraccia, P. Finelli,* and G. G. Consalez

*DIBIT, H San Raffaele Scientific Institute, Milan; *Institute of Genetics, University of Bari, Bari; and †DISEM, University of Genova, Genova, Italy*

Received January 16, 1996

Glycogen synthesis is impaired in first degree relatives of subjects with non-insulin-dependent diabetes mellitus and genes relevant to this metabolic pathway are considered reasonable candidates in the pathogenesis of the disease. In skeletal muscle the *de novo* synthesis of glycogen is primed by an enzyme named glycogenin. We have cloned the glycogenin cDNA from human skeletal muscle mRNA: human glycogenin is a 333 amino acid protein exhibiting 93% identity with rabbit glycogenin. A single transcript of about 2.4 kb, prominent in skeletal muscle, was detected by Northern blot analysis. *In situ* hybridization unequivocally located the human glycogenin gene to chromosome 3q25.1. Furthermore, we mapped two intronless glycogenin-related sequences to human chromosomes 12 and 13.

Genetic factors play a major role in the pathogenesis of non insulin-dependent diabetes mellitus (NIDDM) and the search for new potential candidate genes is influenced by the current understanding of the pathophysiology of the disease. An impairment in insulin-stimulated nonoxidative glucose disposal seems to be the cause of the insulin resistance -a feature of full blown NIDDM- already present in subjects at high risk to develop the disease such as the glucose-tolerant first degree relatives of NIDDM patients (1, 2). Because nonoxidative glucose disposal primarily reflects glycogen synthesis, efforts have been made towards the cloning of the genes involved in this metabolic cascade (e.g. glycogen synthase, insulin-stimulated protein kinase, protein phosphatase-1, etc.) and the subsequent screening for mutations in individuals with NIDDM (3, 4, 5). Recently, the mechanisms of glycogen biogenesis in muscle have been further elucidated. Glycogen synthase alone is unable to promote the *de novo* synthesis of glycogen and requires the priming activity of a protein named glycogenin. This self-glucosylating enzyme generates an oligosaccharide primer suited to initiate glycogen synthesis reactions (6, 7). Gycogenin is complexed to the 86 kDa catalytic subunit of glycogen synthase and a decreased glycogenin activity may have negative effects upon glycogen biosynthesis. The molecular cloning of the human glycogenin cDNA represents the first step to verify this hypothesis.

MATERIAL AND METHODS

Cloning of the human muscle glycogenin cDNA. A set of primer based on 5′ (ATG codon: 5′-TTCTAAAATTCAGGTCTCGTGAA-3′) and 3′ UTR (5′-GCAGCACCATGACAGATC-3′) sequences of a human glycogenin (GenBank entry #X79537) were used for a reverse transcriptase-polymerase chain reaction (RT-PCR) experiment from human muscle RNA (Clontech, Palo Alto, CA). The fragment of ~1.4 kb that we obtained was cloned into the EcoRI and Clal sites of the pBluescriptll SK +/− plasmid vector (Stratagene, LaJolla) and sequenced.

Northern blot analysis. A premade Northern blot (Clontech, Palo Alto) that contains approximately 2 mg of poly(A)+ RNA from each of 8 human tissues was hybridized with the full coding glycogenin cDNA, ^{32}P labeled using a random-primed kit (Amersham, Buckinghamshire).

Isolation of yeast artificial chromosomes. We synthetised a set of primers in the 3′ UT region (sense: 5′-TTAGGATAAGAGGTGAGAACTGG-3′; antisense: 5′-TTCTAAAATTCAGGTCTCGTGAAAGGTA-3′) of the human glycogenin gene to generate by PCR a sequence tagged site (STS) for the screening of a total human YAC library (ICI, ~35000 clones averaging 350 kb in size). We isolated two YACs (7CE3 and 3ODG5). YAC clones were grown in 10 ml of AHC medium for 3 days at 30°C and DNA was extracted by standard procedures.

0006-291X/96 $18.00

Chromosomal localization. Physical mapping was achieved by fluorescence *in situ* hybridization (FISH) to human metaphase chromosomes, using both YACs as probes. YACs were amplified with Alu-PCR, the amplification products labeled with biotin by nick translation, and hybridized *in situ*. Hybridization was performed at 37°C in a volume of 10μl. Biotin-labeled DNA was detected with Cy3-conjugated avidin. Chromosome identification was obtained by simultaneous DAPI staining, that produces a Q-banding pattern.

Southern blot analysis. Human genomic and YAC DNAs were digested with restriction enzymes whose sites were absent or present only once in the human glycogenin cDNA and blotted onto a nylon membrane (Amersham, Buckinghamshire); α-[^{32}P]CTP-labeled glycogenin cDNA was used as probe. Hybridizations and washes were done at high stringency conditions.

Somatic cell hybrid panel. The panel used in this study has been previously characterized (8). The chromosomal mapping of glycogenin-related sequences was achieved by PCR using somatic cell hybrid DNAs and primers designed in the coding region of the glycogenin gene (sense: 5′-ACAGTGGCGATTCTGCTCATCTAAC-3′; antisense: 5′-AAACGGCT-GTGCCATAGCTGGGAT-3′).

RESULTS

Cloning of the Human Muscle Glycogenin cDNA

The rabbit skeletal muscle glycogenin cDNA was recently cloned by other investigators (9); while attempting a "degenerate" oligonucleotide approach to obtain the human homologue, we found that a human glycogenin sequence had been submitted to GenBank with entry #X79537. We synthetised a set of primer based on this sequence and used them for a RT-PCR experiment by which we obtained a fragment of $\sim$1.4 kb. Sequencing of our clone revealed an open reading frame starting at base #10 and ending at base #1002 with a TAG stop codon. Our clone (Genbank entry U44131) differs from the deposited human sequence for an "insertion" of 220 nt in the middle of the coding region. The deduced peptide sequence of our clone gives rise to 333 aminoacid protein displaying *93% identity* with rabbit skeletal muscle glycogenin (Fig. 1).

Northern Blot Analysis

The analysis of mRNA levels in human tissues showed a single transcript of about 2.4 kb prominent in skeletal muscle and heart and present to a lesser extent in lung, kidney, brain, pancreas and placenta (Fig. 2); a very weak signal was detected in liver after an 8 day exposure.

Chromosomal localization. Interspersed repetitive sequence PCR (IRS-PCR) of YACs 7CE3 and 3ODG5 were utilized as probes for FISH analysis; hybridization gave rise to a clear signal on chromosome 3q25.1 (Fig.3).

Southern Blot Analysis.

Southern blot analysis of human genomic DNA suggested a complex organization of the glycogenin gene and/or the presence of glycogenin-related sequences (not shown). To exclude the possibility that the STS could have been located in a glycogenin-related sequence we hybridized with glycogenin cDNA EcoR I-digested DNAs from the YAC clones and from three normal individuals. The comparison revealed that three out of five bands visible in genomic DNA (12, 5 and 2.6 kb) are present in DNAs from YACs 7CE3 and 3ODG5 (Fig. 4). The top band was found to be polymorphic, showing a distinct fragment 23 kb in YAC DNA. Two bands of about 6.3 and 2.3 kb were only visible in genomic DNA.

Somatic Cell Hybrid Panel/Chromosomal Mapping

Previous PCRs from human genomic DNA resulted in the amplification of products colinear with cDNA, a finding suggestive of the existence of intronless glycogenin-related sequence(s). A PCR performed with two primers located in the coding sequence of the glycogenin gene produced a 700 bp fragment -colinear with glycogenin cDNA- in seven (out of 17) somatic hybrid DNAs and in two human genomic samples. All PCR products were digested with either EcoR V or Nco I; as shown in Fig. 5 the amplicons obtained from human templates contain two different DNA species

```
      0                  20                  40                  60                  80
gcagcaccATGACAGATCAGGCCTTTGTGACACTAACCACAAACGATGCCTACGCCAAAGGTGCCCTGGTCCTGGGATCATCTCTGAAACAGCACAGGAC
     MetThrAspGlnAlaPheValThrLeuThrThrAsnAspAlaTyrAlaLysGlyAlaLeuValLeuGlySerSerLeuLysGlnHisArgTh

    100                 120                 140                 160                 180
CACCAGGAGGCTGGTCGTGCTCGCCACCCCTCAGGTCTCAGACTCCATGAGAAAAGTTTTAGAGACAGTCTTTGATGAAGTCATCATGGTAGATGTCTTG
rThrArgArgLeuValValLeuAlaThrProGlnValSerAspSerMetArgLysValLeuGluThrValPheAspGluValIleMetValAspValLeu
 Ser         Ala      Thr                 Thr         Ala      Ile                 Thr      Ile
    200                 220                 240                 260                 280
GACAGTGGCGATTCTGCTCATCTAACCTTAATGAAGAGGCCAGAGTTGGGTGTCACGCTGACAAAGCTCCACTGCTGGTCGCTTACACAGTATTCAAAAT
AspSerGlyAspSerAlaHisLeuThrLeuMetLysArgProGluLeuGlyValThrLeuThrLysLeuHisCysTrpSerLeuThrGlnTyrSerLysC

    300                 320                 340                 360                 380
GTGTATTCATGGATGCAGATACTCTGGTCCTAGCAAATATTGATGATCTTTTTGACAGAGAAGAATTGTCAGCAGCACCAGACCCAGGGTGGCCTGACTG
ysValPheMetAspAlaAspThrLeuValLeuAlaAsnIleAspAspLeuPheAspArgGluGluLeuSerAlaAlaProAspProGlyTrpProAspCy

    400                 420                 440                 460                 480
CTTCAATTCCGGAGTCTTCGTTTATCAGCCTTCAGTTGAAACATACAATCAGCTGTTGCATCTTGCTTCTGAGCAAGGTAGTTTTGATGGTGGGGACCAA
sPheAsnSerGlyValPheValTyrGlnProSerValGluThrTyrAsnGlnLeuLeuHisLeuAlaSerGluGlnGlySerPheAspGlyGlyAspGln
                                                                    Val
    500                 520                 540                 560                 580
GGCATACTGAACACATTTTTTAGCAGCTGGGCAACAACAGATATCAGAAAACACCTGCCGTTTATTTATAACCTAAGCAGCATCTCTATATACTCCTACC
GlyIleLeuAsnThrPhePheSerSerTrpAlaThrThrAspIleArgLysHisLeuProPheIleTyrAsnLeuSerSerIleSerIleTyrSerTyrL
   Leu              Asn
    600                 620                 640                 660                 680
TCCCGGCATTTAAAGTGTTTGGTGCAAGTGCCAAAGTTGTGCATTTCCTGGGACGAGTCAAACCATGGAATTATACTTATGATCCCAAAACAAAAAGTGT
euProAlaPheLysValPheGlyAlaSerAlaLysValValHisPheLeuGlyArgValLysProTrpAsnTyrThrTyrAspProLysThrLysSerVa
            Ala         Asn                         GlnThr                          Thr
    700                 720                 740                 760                 780
CAAAAGTGAGGCCCATGATCCCAACATGACTCATCCAGAGTTTCTCATCCTGTGGTGGAACATCTTTACCACCAACGTTTTACCTCTGCTTCAACAATTT
lLysSerGluAlaHisAspProAsnMetThrHisProGluPheLeuIleLeuTrpTrpAsnIlePheThrThrAsnValLeuProLeuLeuGlnGlnPhe
 Arg      Gly      Thr                 Gln      AsnVal      Asp             Ser   Val
    800                 820                 840                 860                 880
GGCCTTGTCAAAGACACCTGCTCATATGTAAATGTGGAAGATGTCTCAGGAGCCATATCACATCTGTCCCTTGGGGAGATCCCAGCTATGGCACAGCCGT
GlyLeuValLysAspThrCysSerTyrValAsnValGluAspValSerGlyAlaIleSerHisLeuSerLeuGlyGluIleProAlaMetAlaGlnProP
        Gln               GlnHis                      Val                         Thr      ThrThr
    900                 920                 940                 960                 980
TTGTATCCTCGGAAGAACGGAAGGAACGATGGGAACAGGGCCAGGCTGATTATATGGGAGCAGATTCCTTTGACAACATCAAGAGGAAACTTGACACTTA
heValSerSerGluGluArgLysGluArgTrpGluGlnGlyGlnAlaAspTyrMetGlyAlaAspSerPheAspAsnIleLysArgLysLeuAspThrTy
                                                                                    Lys
   1000                1020                1040                1060                1080
CCTCCAGTAGaaacactgcatttttctgtgaacacatccacttcacaagccttgtttctgatacttagtatctagagctgggttgagcaaagtctgttca
rLeuGlnEnd
      Pro
   1100                1120                1140                1160                1180
gttgctagaggttttcattaaaacttatcagatgagaggctttttaggataagaggtgagaactgggcaaaagttgtgaagcagcaattctgttatatg

   1200                1220                1240                1260                1280
gacagtgttctgctttttaatcctatttagcttgtttcagaaattctcacttttgttgactgccaacatacaaagtaagggaaactcaagatattaagat

   1300                1320                1340                1360                1380
ggctgtatcagttcttaaaatctgcagagcctggttcaaaatcagtcactcccttcagaagcagacatggcatctgttccttgcttgcttgttggttgtg

   1400                1420
tacctttcacgagacctgaattttagaa
```

FIG. 1. Nucleotide sequence of human skeletal muscle glycogenin, its deduced amino acid sequence and comparison with rabbit skeletal muscle glycogenin.

that can be distinguished by the presence/absence of EcoR V and Nco I sites. Four somatic cell hybrids, which contain chromosome 13, bear the species in which both EcoR V and Nco I sites are absent. Differently, the three hybrids that carry the fragment digested by EcoR V and Nco I share human chromosome 12. The same pair of primers did not amplify the 700 bp fragment from the two YACs used for FISH analysis.

The primer pair designed for YAC library screening (i.e. STS) amplified a fragment of the appropriate size (280 bp) from DNA of the four hybrids containing human chromosome 3 (not shown).

DISCUSSION

In the present study, we have cloned and sequenced the coding cDNA of the human muscle

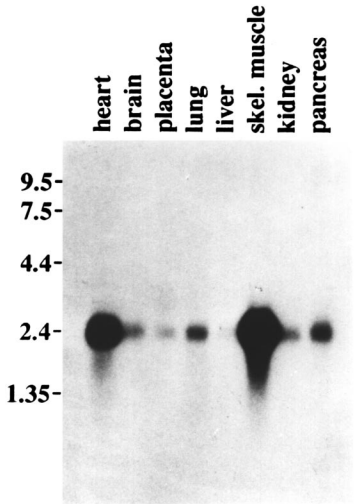

FIG. 2. Northern blot analysis of the human glycogenin gene; 48 hr exposure. A very faint band is detected in liver (not visible here) after an 8 day exposure.

glycogenin: the open reading frame of our clone yields a 333 aminoacid protein which displays 93% identity with rabbit glycogenin. In contrast, the deduced amino acid sequence of the other human clone exhibits a substantial identity with rabbit glycogenin up to aa. #202 but a totally different sequence in the rest of the protein.

By Northern blot we found a single transcript of about 2.4 kb, even at low stringency conditions: this result supports the notion that only one glycogenin gene is expressed in the tissues examined.

FIG. 3. FISH of a human metaphase using labeled IRS-PCR products from YAC 30DG5 as a probe. The YAC contains the glycogenin gene. The green arrows indicate specific labeling at chromosome 3q25.1.

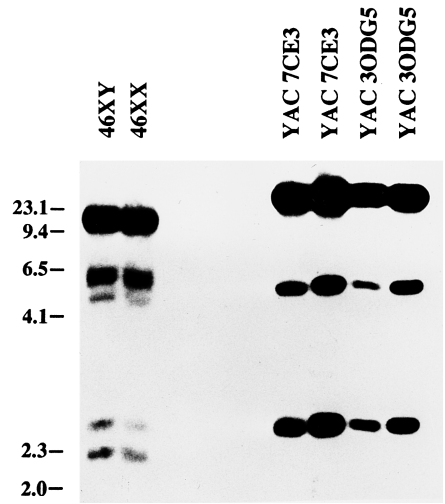

FIG. 4. Southern blot of EcoR I-digested human genomic DNA and of DNA from YAC 7CE3 and 30DG5, hybridized with glycogenin cDNA. Note the two extra bands of about 6.3 and 2.3 kb detected in human genomic DNA.

The pattern of expression in human tissues is superimposable to that found in rabbit (9), in particular the lowest levels of glycogenin mRNA is detected in liver, an organ with large deposits of glycogen. In rabbit, this low abundance corresponds to the modest amounts of glycogenin protein found in this organ. Liver glycogenin in rabbit is indistinguishable from muscle glycogenin by amino acid sequencing (6) and likely represents the same gene product. According to our results, this interpretation could be extended to humans. However, given the differences between the liver and skeletal muscle isoforms of glycogen synthase (10) and the fact that glycogenin is supposed to have a tight interaction with this enzyme, the existence of a liver glycogenin isoform cannot be excluded.

Our work has assigned the human glycogenin gene to chromosome 3q25.1. PCR-based analysis of the somatic cell hybrid panel has confirmed that the glycogenin gene is located on chromosome 3 and have mapped two glycogenin-related sequences to chromosome 13 and 12. The latter likely represent retrotransposed pseudogenes: the two fragments of ~6.3 and 2.3 kb detected by Southern blot in human genomic (but not YAC) DNA probably contain these glycogenin-like sequences. Noteworthy, the existence of glycogenin pseudogenes in the rabbit genome has been proposed previously (9).

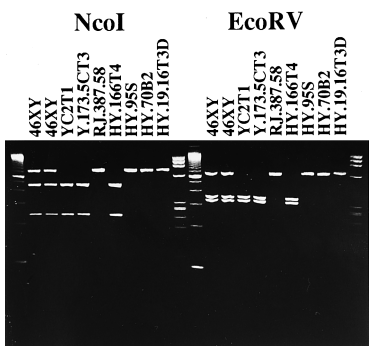

FIG. 5. The 700 bp PCR amplification fragment was digested with either EcoR V or Nco I revealing two different amplicons in human genomic DNA. The two products were assigned to chromosome 12 (digested fragment) and 13 (uncut fragment). HY.166T4, Y.173.5CT3, YC2T1 share chromosome 12; HY.19.16T3D, HY.70B2, HY.95S and RJ.387.58 share chromosome 13.

The cDNA and physical mapping of human glycogenin gene represent new tools towards a better understanding of the mechanisms of glycogen synthesis and -possibly- the pathophysiology of insulin resistance of NIDDM.

ACKNOWLEDGMENTS

The financial support of Telethon-Italy (Grants E.228 and E.098) is gratefully acknowledged. G.G.C. is supported by Grant B14 from Telethon-Italy. Thanks to C. Sala and D. Toniolo for the isolation of YAC clones. The YAC Screening Centre (DIBIT-HSR, Milan, Italy) is supported by Grant C08 from Telethon-Italy. Part of these studies have been presented at the 31st Annual Meeting of the European Association for the Study of Diabetes (EASD), Stockholm, Sweden, Sept. 1995 (*Diabetologia* 1995, **38**) (Suppl. 1). Abstract 295.

REFERENCES

1. Vaag, A., Henriksen, J. E., and Beck-Nielsen, H. (1992) *J Clin Invest* **89,** 782–788.
2. Gulli, G., Ferrannini, E., Stern, M., Haffner, S., and DeFronzo, R. A. (1992) *Diabetes* **41,** 1575–1586.
3. Bjorbaek, C., Echwald, S., Hubricht, P., Vestergaard, H., Hansen, T., Zierath, J., and Pedersen, O. (1994) *Diabetes* **43,** 976–983.
4. Bjornbaek, C., Vik, T. A., Echwald, S. M., Yang, P. Y., Vestergaard, H., Wang, J. P., Webb, G. C., Richmond, K., Hansen, T., Erikson, R. L., Gabor Miklos, G. L., Cohen, P. T. W., and Pedersen, O. (1995) *Diabetes* **44,** 90–97.
5. Chen, Y. H., Hansen, L., Chen, M. X., Bjorbaek, C., Vestergaard, H., Hansen, T., Cohen, P. T. W., and Pedersen, O. (1994) *Diabetes* **43,** 1234–1241.
6. Smythe, C., Villar-Palasi, C., and Cohen, P. (1989) *Eur J Biochem* **183,** 205–209.
7. Smythe, C., and Cohen, P. (1991) *Eur J Biochem* **200,** 625–631.
8. Antonacci, R., Marzella, R., Finelli, P., Lonoce, A., Forabosco, A., Archidiacono, N., and Rocchi, M. (1995) *Cytogenet Cell Genet* **68,** 25–32.
9. Viskupic, E., Cao, Y., Zhang, W., DePaoli-Roach, A. A., and Roach, P. J. (1992) *J Biol Chem* **267,** 25759–25763.
10. Nuttal, F. Q., Gannon, M. C., Bai, G., and Lee, E. Y. (1994) *Arch Biochem Biophys* **311,** 443–449.