Identification of the Human Methylmalonyl-CoA Racemase Gene Based on the Analysis of Prokaryotic Gene Arrangements

IMPLICATIONS FOR DECODING THE HUMAN GENOME*

Thomas A. Bobik† and Madeline E. Rasche
From the Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611

In this report, we identify the human DL-methylmalonyl-CoA racemase gene by analyzing prokaryotic gene arrangements and extrapolating the information obtained to human genes by homology searches. Sequence similarity searches were used to identify two groups of homologues that were frequently arranged with prokaryotic methylmalonyl-CoA mutase genes, and that were of unknown function. Both gene groups had homologues in the human genome. Because methylmalonyl-CoA mutases are involved in the metabolism of propionyl-CoA, we inferred that conserved neighbors of methylmalonyl-CoA mutase genes and their human homologues were also involved in this process. Subsequent biochemical studies confirmed this inference by showing that the prokaryotic gene PH0272 and its human homologue both encode DL-methylmalonyl-CoA racemases. To our knowledge this is the first report in which the function of a eukaryotic gene was determined based on the analysis of prokaryotic gene arrangements. Importantly, such analyses are rapid and may be generally applicable for the identification of human genes that lack homologues of known function or that have been misidentified on the basis of sequence similarity searches.

Prokaryotes frequently cluster genes of related function. Hence, if the function of one gene in a conserved cluster is known, the remaining genes in that cluster can be inferred to function in the same metabolic process (1–5). Knowledge of a gene’s metabolic role can lead to rapid assignment of specific function by focusing biochemical studies. Although such analyses appear inapplicable to eukaryotes (gene clustering is rare in these organisms), it should be possible to extrapolate the functional information obtained from the analysis of prokaryotic gene arrangements to eukaryotes by homology searches (Fig. 1). To determine whether such analysis can in fact be used to rapidly determine the function of human genes, we investigated genes involved in propionyl-CoA metabolism.

Both prokaryotes and eukaryotes metabolize propionyl-CoA by a coenzyme B12-dependent pathway (6) (Fig. 2). In humans, inborn errors in the methylmalonyl-CoA mutase (mcm) gene lead to methylmalonyl aciduria, a rare but severe inherited disease (7). Defects in the CblABCDF complementation groups also lead to this disease (7). These complementation groups are thought to encode enzymes needed for the conversion of hydroxy-B12 to coenzyme B12, the required cofactor for methylmalonyl-CoA mutase (MCM). In addition, defects in the DL-methylmalonyl-CoA racemase gene may also lead to methylmalonyl aciduria, as MCM is specific for the l-isomer of methylmalonyl-CoA (8). Although the mcm gene is known, the genes corresponding to the CblABCDF complementation groups and to the DL-methylmalonyl-CoA racemase have not been identified.

To identify additional human genes involved in propionyl-CoA metabolism, prokaryotic gene arrangements were examined. By these analyses, we identified the human DL-methylmalonyl-CoA racemase gene as well as a second human gene likely to be involved in propionyl-CoA metabolism, but of unknown function. To our knowledge, this is the first report of a eukaryotic gene that has been identified based on the analysis prokaryotic gene arrangements. Importantly, the method used is rapid and may be generally applicable for the identification of eukaryotic genes that lack homologues of known function or which have been misidentified on the basis of sequence similarity searches.

MATERIALS AND METHODS

General Protein and Molecular Methods—Electrophoresis, bacterial transformation, restriction digestions, and other routine molecular and biochemical methods were performed as described previously (9).

Cloning of the Human in-Methylmalonyl-CoA Racemase Coding Sequence for High-Fidelity Expression—PCR was used to amplify the portion of the human racemase cDNA corresponding to its prokaryotic homologues. The primers used were GTGTGGAACCTGGGTCGACTGACCA and CTGGATGACGACGACAAGATGTGGAACCTGGGTCGA, and the template used was a Marathon-ready human liver cDNA library (CLONTECH, Palo Alto, CA). The enzyme used for amplification was Advantage 2 DNA polymerase mix (CLONTECH). The PCR product obtained was treated with the Klenow fragment of DNA polymerase to create blunt-ended DNA (10) and cloned into the expression vector, pET41a (Novagen, Cambridge, MA). The BglII site was positioned such that after cloning the in-methylmalonyl-CoA racemase would be fused to N-terminal glutathione S-transferase- and 6×His-tags encoded by the expression vector. The enterokinase site was posi

* This work was supported in part by Grant GM59486 (to T. A. B.) from the National Institutes of Health and Grant MCB-9876212 (to M. E. R.) from the National Science Foundation. Florida Agricultural Experiment Station Journal Series No. R-08334. The costs of publication are chargeable. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence and requests for materials should be addressed: Dept. of Microbiology and Cell Science, Bldg. 981, Rm. 1220, University of Florida, Gainesville, FL 32611 Tel.: 352-846-0957; Fax: 352-392-5922; E-mail: bobik@ufl.edu.

‡ To whom correspondence and requests for materials should be addressed: Dept. of Microbiology and Cell Science, Bldg. 981, Rm. 1220, University of Florida, Gainesville, FL 32611 Tel.: 352-846-0957; Fax: 352-392-5922; E-mail: bobik@ufl.edu.
Identification of the Human Methylmalonyl-CoA Racemase

**Growth of Expression Strains and Preparation of Cell Extracts**

**FIG. 2.** Coenzyme B12-dependent propionyl-CoA metabolism. Propionyl-CoA carboxylase catalyzes formation of the t-isomer of methylmalonyl-CoA. mC-Methylmalonyl-CoA racemase is needed to catalyze the conversion of t-methylmalonyl-CoA to L-methylmalonyl-CoA. The enzyme B12-dependent methylmalonyl-CoA racemase catalyzes the conversion of t-methylmalonyl-CoA to L-methylmalonyl-CoA. The enzyme converts D-methylmalonyl-CoA to L-methylmalonyl-CoA, which is then converted to succinyl-CoA by MCM, which is specific for the L-isomer (6). Racemase activity was quantified using HPLC to follow the disappearance of methylmalonyl-CoA using HPLC conditions described previously (11). Assay mixtures contained 50 mM potassium phosphate pH 7, 25 mM NaCl, 2 mM MgCl2, 75 mM L-methylmalonyl-CoA, and 5.6 \( \mu \)g/ml holo-MCM. The holo-MCM was prepared by incubating the following mixture for 30 min at 4 °C in the dark: 0.56 \( \mu \)g/\( \mu \)l purified apo-MCM, 0.63 \( \mu \)g coenzyme B12, and 2 mM dithiothreitol. The specific activity of the holo-MCM used was >2 \( \mu \)mol/min/mg of protein. Assays were initiated by addition of holo-MCM and incubated for 5 min at 37 °C. This depleted the L-methylmalonyl-CoA from the assay mixtures. After this initial 5-min incubation, a source of t-methylmalonyl-CoA racemase was added, and incubation at 37 °C was continued for an additional 5 min. Reactions were terminated by addition of 75 \( \mu \)l of 1 M acetic acid. Samples were then frozen using a dry ice ethanol bath and stored at −20 °C until analyzed. To prevent photolyzing of coenzyme B12, all manipulations were carried out in dim light.

**Accession numbers**—The GenBank accession number for Homo sapiens methylmalonyl-CoA racemase cDNA is AF364547. The GenBank accession number for P. horikoshii methylmalonyl-CoA racemase coding sequence is AF364548.

**RESULTS**

Identification of Conserved Neighbors of Prokaryotic mcm genes—Prokaryotes often cluster genes involved in the same metabolic process (1–5). Therefore, to identify additional genes involved in propionyl-CoA metabolism, we searched for genes that frequently cluster with propionyl-CoA genes (Fig. 1). The first step employed was to identify propionyl-CoA genes that had both mcm homologues and completed genome sequences. Blast searches (12) of the NCBI nonredundant (nr) database identified 16 mcm genes that had homologues (Excluded: \( 7 \times 10^{-11} \)) of the human mcm gene. For eight of these, complete genome sequences were available. These included three Bacteria (Deinococcus radiodurans, Mycobacterium tuberculosis, and E. coli) and five Archaea (Archaeoglobus fulgidus, *P. horikoshii*, Pyrococcus abyssi, Aeropyrum pernix, and Halobacterium sp. NRC-1). Next, the chromosomal contexts of the mcm genes were examined.
Identification of the Human Methylmalonyl-CoA Racemase

In these eight prokaryotes, Blastp software was used to identify the homologues of ten genes on each side of the mcm gene. Then, it was determined whether these homologues also clustered with mcm genes by examining the protein tables available through Entrez Genomes (www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html). By these analyses, two conserved neighbors of prokaryotic mcm genes were identified: genes with homology to lactoylglutathione (LGSH) lyases and LAO transporters were neighbors of mcm genes in 5/8 prokaryotic genomes and in 4/5 Archaeal genomes examined, and genes with homology to lysine/arginine/ornithine (LAO) transport proteins were found near mcm homologues in 8/8 cases. These gene arrangements suggested that certain proteins with homology to LGSH lyases and LAO transporters were misidentified by homology searches and actually encode proteins involved in propionyl-CoA metabolism.

Identification of the Human Homologues of Conserved Neighbors of Prokaryotic mcm Genes—Next we sought to identify the human homologues of prokaryotic LGSH lysases and LAO transporters. The P. horikoshii protein gi7448629 (an LGSH lyase homologue) was used to query the human genome with the homology searches employed and in fact encoded a protein involved in propionyl-CoA metabolism.

The full-length human cDNA identified above encoded a protein with about 40 N-terminal amino acids not found in its prokaryotic homologues (Fig. 3). In humans, propionyl-CoA metabolism occurs in the mitochondrion (7), and the additional amino acids of the protein encoded by this cDNA aligned with the prokaryotic sequence used in the purification of this enzyme.

To identify the human homologue of prokaryotic LAO transport proteins, a procedure similar to that described above for identification of the human LGSH lyase homologue was employed. The longest human cDNA identified (gi: 12914481) was a partial sequence that represented four exons on chromosome IV. The protein encoded by this cDNA aligned with the M. tuberculosis LAO transporter (gi: 3915555) over the greater portion of its length (263/334) with an Expect value of $3 \times 10^{-30}$ using Blastx software. Because the human cDNA was incomplete at the 5’ end, the possibility of a mitochondrial targeting sequence could not be examined. Nonetheless, the cDNA identified represents a probable human LAO transporter homologue.

Thus, two human genes were identified that were homologous to conserved neighbors of prokaryotic mcm genes. We inferred that these human genes were involved in propionyl-CoA metabolism. Unidentified genes thought to function in this process include those that encode 3-methylmalonyl-CoA racemases as well as those that encode enzymes needed to convert hydroxy-B12 to coenzyme B12. Accordingly, we focused further studies on determining whether the human cDNAs identified encoded such enzymes. A 3-methylmalonyl-CoA racemase was previously purified from rat liver and was found to be composed
Identification of the Human Methylmalonyl-CoA Racemase

The predicted human racemase was then purified using several affinity chromatography steps as well as enterokinase cleavage of the racemase fusion protein. The site of enterokinase cleavage is shown in Fig. 3. The progress of the purification was monitored by SDS-polyacrylamide gel electrophoresis and coenzyme B12 are required components, because the racemase reaction involves coupled enzymatic reactions.

Purification of the Human DL-Methylmalonyl-CoA Racemase—The predicted human racemase was then purified using several affinity chromatography steps as well as enterokinase cleavage of the racemase fusion protein. The site of enterokinase cleavage is shown in Fig. 3. The progress of the purification was monitored by SDS-polyacrylamide gel electrophoresis followed by Coomassie staining (Fig. 4). The specific activity of the racemase increased at each step of the purification (Table I). The final specific activity of the highly purified racemase was 833 μmol/min/mg of protein. This activity is within the range previously reported for DL-methylmalonyl-CoA racemases purified from other sources. DL-Methylmalonyl-CoA racemase previously purified from Propionibacterium shermanii had a specific activity of 33.4 or 607.5 μmol/min/mg of protein depending on the purification protocol used (17, 18), and DL-methylmalonyl-CoA racemase purified from rat liver had a specific activity of 8400 μmol/min/mg of protein (16). Thus, the findings presented above constitute strong evidence that one of the two human genes identified as a conserved neighbor of prokaryotic mcm genes encodes a DL-methylmalonyl-CoA racemase rather than an LGSH lyase as would be suggested by homology searches.

Expression of the P. horikoshii DL-Methylmalonyl-CoA Racemase—To obtain further evidence that certain LGSH lyase homologues are in fact DL-methylmalonyl-CoA racemases, the predicted P. horikoshii racemase gene was cloned into a T7 expression vector. This clone mediated expression of a large amount of protein near the expected molecular mass, 15 kDa (not shown). Partially purified cell extracts were prepared by anion exchange chromatography, and those derived from cells containing the expression plasmid with insert (strain TA1015) contained racemase enzyme with a specific activity of 93 μmol/min/mg of protein. In contrast, no racemase activity was observed in cell extracts derived from cells containing vector without insert (strain BE119). This indicated that the cloned P. horikoshii gene encoded a DL-methylmalonyl-CoA racemase. These findings provided evidence that LGSH lyase homologues neighboring mcm genes were misidentified by homology searches and actually encode DL-methylmalonyl-CoA racemases. In addition, they provide confirmatory evidence that the human clone described above also encodes a DL-methylmalonyl-CoA racemase.

DISCUSSION

In this report, we sought to determine whether the analysis of prokaryotic gene arrangements could be used to help determine the function of human genes. Sequence similarity searching identified two genes of unknown function that were conserved neighbors of prokaryotic mcm genes and that had human homologues. Because MCMs are involved in propionyl-CoA metabolism, we inferred that conserved neighbors of mcm genes and their human homologs were also involved in this process, and we used this knowledge to guide biochemical tests. Subsequently, we showed that one conserved neighbor (PHO0272) and its human homologue both encoded DL-methylmalonyl-CoA racemases. The purified human racemase had a
specific activity of 833 μmol/min/mg of protein, which is comparable with the specific activity of DL-methylmalonyl-CoA racemases purified from other sources (see above) (16–18). These findings constitute strong evidence that the gene corresponding to cDNA af364547 is indeed a DL-methylmalonyl-CoA racemase gene. Thus, a previously unknown human gene was rapidly identified based on the analysis of prokaryotic gene arrangements.

The human DL-methylmalonyl-CoA racemase gene has been implicated in the inherited disorder methylmalonic aciduria (although this is uncertain, since a bypass pathway of D-methylmalonyl-CoA racemization may exist in mammals) (8). Thus, the analysis of prokaryotic gene arrangements has the potential to allow the rapid identification of human genes involved in inherited metabolic disorders. Importantly, the prediction that a given gene is involved in a particular inherited disease does not require the determination of its biochemical function. Such predictions require only knowledge that a gene is involved in a particular metabolic process, and this information can be obtained directly from the bioinformatic analysis of gene arrangements.

We also identified a second gene of unknown function that was a conserved neighbor of mcm genes and that had homology to a human gene. This gene had homology to those that encode LAO transport proteins. The function of this gene has not yet been determined. We suspect that it is not involved in lysine/arginine/ornithine transport. It may have a role in the conversion of hydroxy-B12 to coenzyme B12, or in the reactivation of MCM, as genes encoding such enzymes have been shown to cluster with genes encoding other coenzyme B12-dependent enzymes (9, 13, 14, 19). Attempts to determine the function of this gene with biochemical tests were hampered by the fact that overexpression resulted in inclusion body formation under all conditions tested. Nonetheless, the potential involvement of this gene in methylmalonyl aciduria is testable given its DNA sequence even without knowledge of its specific biochemical function.

It was expected that several genes involved in the conversion of vitamin B12 to hydroxy-B12 would be identified by the gene arrangement analyses reported herein. As mentioned above, only one candidate was found and its function was not verified. There are several possible reasons why the genes for such enzymes were not found. We feel the most likely reason is that some of the genes involved in the conversion of hydroxy-B12 to coenzyme B12 have functions in addition to propionyl-CoA metabolism that dictate alternative gene arrangements.

Last, we point out that the analysis of prokaryotic gene arrangements may have broad application for the identification of human genes. Potentially, the method is applicable to any physiological process that is shared between prokaryotes and eukaryotes. This includes many aspects of metabolism, transcription, translation, DNA replication, and DNA repair. Furthermore, the method may prove valuable for the identification of genes involved processes currently thought to be distinctively eukaryotic.

Acknowledgments—We thank the National Institutes of Health, the National Science Foundation, the Florida Agricultural Experiment Station, Peter E. Kima, Nemat O. Keyhani, and Roy A. Jensen.

REFERENCES
1. Bork, P., Dandekar, T., Diaz-Lazaro, Y., Eisenhaber, F., Huynen, M., and Yuan, Y. (1998) J. Mol. Biol. 283, 707–725
2. Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Trends Biochem. Sci. 23, 324–328
3. Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) Nature 405, 823–826
4. Marcotte, E. M. (2000) Curr. Opin. Struct. Biol. 10, 359–365
5. Overbeek, R., Fonstein, M., D’Souza, M., Posch, G. D., and Metzal, N. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2896–2901
6. Banerjee, R., and Chowdhury, S. (1999) in Chemistry and Biochemistry of B12 (Banerjee, J., ed) pp. 707–730, John Wiley, New York
7. Rosenblatt, D. S., and Fenton, W. A. (1999) in Chemistry and Biochemistry of B12 (Banerjee, R., ed) pp. 367–384, John Wiley, New York
8. Montgomery, J. A., Mamer, O. A., and Sriver, C. R. (1983) J. Clin. Invest. 72, 1957–1947
9. Johnson, C. L. V. J., Pechonick, E., Park, S. D., Havemann, G. D., Havemann, G. D., and Bobik, T. A. (2001) J. Bacteriol. 183, 1577–1584
10. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
11. Demoz, A., Garras, A., Asiedu, D. K., Netteland, B., and Berge, R. K. (1995) J. Chromatogr. 667, 148–152
12. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
13. Toraya, T., and Mori, K. (1999) J. Biol. Chem. 274, 3372–3377
14. Seifert, C., Bowien, S., Gottschalk, G., and Daniel, R. (2001) Eur. J. Biochem. 268, 2369–2378
15. Claros, M. G., and Vincens, P. (1996) Eur. J. Biochem. 241, 770–786
16. Stabler, S. P., Marcell, P. D., and Allen, R. H. (1985) Arch. Biochem. Biophys. 237, 262–264
17. Allen, S. G. H., Kellermeyer, R., Stjernholm, R., Jacobson, B., and Wood, H. G. (1963) J. Biol. Chem. 238, 1637–1642
18. Leadlay, P. F. (1983) Biochem. J. 197, 413–419
19. Mori, K., Tohimitsu, T., Hara, T., and Toraya, T. (1997) J. Biol. Chem. 272, 32034–32041

---

2. T. A. Bobik and M. E. Rasche, unpublished data.
Identification of the Human Methylmalonyl-CoA Racemase Gene Based on the Analysis of Prokaryotic Gene Arrangements: IMPLICATIONS FOR DECODING THE HUMAN GENOME

Thomas A. Bobik and Madeline E. Rasche

J. Biol. Chem. 2001, 276:37194-37198.
doi: 10.1074/jbc.M107232200 originally published online July 31, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M107232200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 16 references, 6 of which can be accessed free at http://www.jbc.org/content/276/40/37194.full.html#ref-list-1