European Journal of Human Genetics (2000) 8, 367–371

www.nature.com/ejhg

ARTICLE

Lack of Hardy-Weinberg equilibrium for the most prevalent *PMM2* mutation in CDG-Ia (congenital disorders of glycosylation type Ia)

Els Schollen[1], Susanne Kjaergaard[2], Eric Legius[1], Marianne Schwartz[2] and Gert Matthijs[1]

[1]*Center for Human Genetics, University of Leuven, Belgium;* [2]*Department of Clinical Genetics, Rigshospitalet, Copenhagen, Denmark*

The R141H mutation in the *PMM2* gene is the most frequent mutation in type Ia of the congenital disorders of glycosylation (formerly carbohydrate-deficient glycoprotein syndromes)(CDG-Ia). However, it has never been observed in the homozygous state. Homozygosity for this mutation is probably incompatible with life. In this study, we determined the frequency of R141H in two normal populations: in neonates of Dutch origin 1/79 were carriers, whilst in the Danish population, a carrier frequency of 1/60 was found. These figures are clearly in disequilibrium with the frequency of CDG-Ia that has been estimated at 1/80 000 to 1/40 000 in these populations. Haplotype analysis of 43 patients with the R141H mutation of different geographic origins indicated that the R141H is an old mutation in the Caucasian population. Based on the new data, the disease frequency has been calculated at 1/20 000 in these populations. It is concluded that the disease is probably underdiagnosed. *European Journal of Human Genetics* (2000) **8**, 367–371.

Keywords: N-glycosylation; carrier frequency; heterozygote advantage; phosphomannomutase; foetal wastage; recessive mutation

Introduction

Congenital disorders of glycosylation type Ia (CDG-Ia, formerly called carbohydrate-deficient glycoprotein syndrome type I or Jaeken syndrome, OMIM 212065) is an autosomal recessive disorder characterised by a defective glycosylation. Most patients show a deficiency of phosphomannomutase (PMM), the enzyme that converts mannose 6-phosphate to mannose 1-phosphate in the synthesis of GDP-mannose. This subtype of the disease (CDG-Ia) is linked to chromosome 16p13.[1,2] We have previously cloned the *PMM2* gene (OMIM 601785) and identified mutations in patients with a PMM deficiency.[3] In two independent surveys on patients with CDG-Ia, mutations in the coding region of the *PMM2* gene were found on 145 of 148 disease chromosomes (98%).[3–5] The majority of the 30 different mutations detected so far have been of the missense type.[3–7] Probably, only such mutations that retain residual enzymatic activity are tolerated in patients. Extensive allelic heterogeneity has hampered the study of genotype–phenotype correlation.[4,8]

The most intriguing observation is the total lack of patients homozygous for the most frequent mutation, R141H, present in more than 75% of patients of Caucasian origin.[4,5] On the other hand, patients homozygous for the other relatively frequent mutation, F119L, have been found,[4,6,9] (see also M Schwartz 1999, unpublished data) as well as one patient homozygous for the rare D65Y mutation.[4] These observations suggested that the R141H mutation is a severe mutation, and that homozygosity may not be compatible with life.

We determined the frequency of R141H in two normal populations and analysed the haplotypes of the R141H chromosomes in a large series of patients.

Correspondence: Gert Matthijs, Center for Human Genetics, Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.
Tel: +32 16 346070; Fax: +32 16 346060;
E-mail: gert.matthijs@med.kuleuven.ac.be
Received 25 August 1999; revised 4 January 2000; accepted 13 January 2000

Material and methods

Samples

DNA was extracted from 950 Guthrie cards from neonates of Dutch origin and from 420 blood samples obtained from

anonymous donors at the Copenhagen Blood Bank. DNA samples were also available from 600 individuals with mental retardation and a series of 100 couples with at least one miscarriage. These samples had previously been collected in the context of two independent studies on the genetic basis of mental retardation and (recurrent) miscarriage, respectively.

Oligo ligation assay

An oligonucleotide ligation assay (OLA) in a 96-well format was developed for the rapid detection of the R141H and F119L mutations.[10] Exon 5 of *PMM2* was amplified as described,[4] and the PCR products were subsequently used in the OLA with the following primer combinations: for R141H: 5'-biotin-GAAGCTGCAGCCAAGAAGAAC-3', 5'-PO_4-GCATTGAGTTCTACGAACTC-FITC-3' (wild type specific oligonucleotide) and 5'-PO_4-ACATTGAGTT CTACGAACTC-FITC-3' (mutant); for F119L: 5'-biotin-TACATTCCCAGGGGTACTTT-3', 5'- PO_4-AATTGAATTCCG AAATGGG-FITC-3' (mutant) and 5'- PO_4-CATTGAA TTCCGAAATGGG-FITC-3' (wild type). The ligation reaction was done with Tsc-ligase (Roche Diagnostics, Brussels, Belgium) according to the manufacturer's protocol. For the detection, a colorimetric ELISA system (Roche) with an alkaline phosphate-anti-FITC antibody and BM-Blue substrate was used.

Patients

Mutation data were available from 101 unrelated CDG-Ia patients. Among these were patients from 79 families of different origin (see Table 1) analysed in Leuven,[3,4] (including unpublished cases), and from 22 Danish families, analysed in Copenhagen[5] (including unpublished cases). Twenty-eight patients were of Dutch or Flemish origin, and because the populations are closely related, this group was combined for the frequency calculations.

Haplotype analysis

Forty-three CDG-Ia patients with an R141H mutation, of which DNA from the parents was available, were typed for marker D16S3020[11] by PCR. The alleles were sized on an ABI PRISM 310 (PE Biosystems, Nieuwerkerk a/d IJssel, The Netherlands).

Results

Frequency of R141H in two normal populations

Twelve carriers were identified among the 950 Dutch neonates and seven among the 420 Danish individuals. The estimated carrier frequency is thus 1/79 (95% CI = 1/51–1/182) for the Dutch, and 1/60 (95% CI = 1/35–1/227), for the Danish population (Table 2). These figures are not significantly different, and the pooled data give an estimated 1/72 (95% CI = 1/50–1/130). The entire group was also analysed for the presence of the second most frequent mutation, F119L, which is particularly prevalent in Scandinavia.[5,6,12] No carriers were identified in either set, indicating a much lower frequency of this mutation. On the basis of the disease frequency and allele distribution (see below), an estimated 1/500 to 1/1000 was expected.

Table 1 Number of CDG-Ia patients which are either compound heterozygotes for R141H and another mutation or for two different mutations

Country	*R141H*	*no R141H*
Denmark	19 (13)	3
US	7 (2)	1[a]
Belgium	8 (6)	1
Netherlands	14 (8)	5
Germany	10 (5)	2
France	7 (4)	2
Italy	5 (1)	0
UK	3 (2)	0
Ireland	2 (1)	0
Switzerland	1	0
South America	2 (1)	0
Spain	3	2
Portugal	1	1
Japan	0	2

The number of patients included in the haplotype study is given in parentheses; [a]US patient without the R141H mutation has Asian ancestors.

Table 2 Population data on CDG-Ia

	Dutch/Flemish population	*Danish population*	*Pooled data*
Observed carrier frequency for R141H	1/79 (1/51–1/82)	1/60 (1/35–1/227)	1/72 (1/50–1/130)
Observed proportion of R141H heterozygotes in CDG-Ia patients	22/28	19/22	41/50
Calculated frequency of R141H homozygotes (q^2)	1/25 000	1/14 400	1/20 700
Calculated frequency of combinations of mutations other than R141H (r^2)	1/84 000	1/144 400	1/107 600
Calculated frequency of compound heterozygotes for R141H and another mutation (2qr)	1/23 000	1/23 000	1/23 600
Calculated frequency of the disease (q^2+r^2+2qr)	1/10 500	1/8 500	1/10 000
Expected frequency of the disease (r^2+2qr)	1/18 000	1/20 000	1/20 000

Among our series of 101 patients, the allele frequency by counting for R141H is 82/202 or 40%. If we restrict the analysis to the Dutch/Flemish plus Danish populations, 41/50 patients are compound heterozygous for R141H (see Table 1).

Under the Hardy-Weinberg equilibrium, with p = frequency of the normal alleles, q = frequency of R141H alleles and r = frequency of other CDG-Ia mutations, one can use the carrier frequency for R141H in the population (2pq) and the proportion of R141H compound heterozygotes among the patients (2qr), to estimate r. The essential frequency of R141H homozygotes (q^2) can be calculated and an estimate of the frequency of the disease can be made. For the Dutch/Flemish plus Danish population, q^2 (expected frequency of patients homozygous for R141H) is approximately 1/20 700, r^2 (expected frequency of patients with mutations other than R141H) is 1/107 600 and the calculated number of compound heterozygotes (R141H/other mutation) is 1/23 600. Together, this results in an expected frequency of the disease of 1/10 000. Because patients homozygous for R141H have not been found, the disease frequency ($r^2 + 2qr$) is estimated at 1/20 000. Individual calculations for the Dutch/Flemish and Danish populations are given in Table 2.

Second, if p, q and r have been estimated and the frequency of the different genotypes in affected patients ($r^2 + 2qr + q^2$) has been derived from these figures, then the expected proportion of q^2 (R141H homozygotes) can be estimated ($q^2/r^2 + 2qr + q^2$). For the Dutch/Flemish plus Danish population, 25 homozygotes are expected and 0 are observed ($\chi^2 = 50$, df = 1, $P < 10^{-12}$). The number of homozygotes for R141H is thus significantly lower than expected under the Hardy-Weinberg equilibrium. We do not think that this can be accounted for by mutations, genetic drift, non-random mating or migration. It must result from (prenatal) selection.

Frequency of R141H in two selected groups

We have tried to find conditions that might be linked to homozygosity for R141H. We tested the carrier frequency in two selected groups. First, a group of 600 individuals with mental retardation as the princeps symptom was analysed for the presence of R141H, because it might be that the defect presents solely with a neurological picture. No significant increase of the carrier frequency was detected (1/60). Second, the mutation might play a role in foetal wastage and (early) miscarriages. A series of 100 couples with at least one miscarriage was tested. The frequency of the mutation in this group was 2/200, which is not different from the control populations. Thus far we have not identified any couple in which both partners are carriers of R141H.

Haplotype analysis

To investigate the origin of the R141H mutation, patients from different geographical origin and heterozygous for R141H were genotyped for marker D16S3020, which is located within 20 kb of *PMM2*. In 43 patients and their parents (Table 2), the phase of the alleles could be determined. Table 3a gives the distribution of the different alleles on the R141H chromosomes and on the normal chromosomes in the parents. Out of 43 patients, 41 carried the allele '16' for marker D16S3020; this allele was found only on nine of the 86 normal chromosomes ($\chi^2 = 83$, $P < 10^{-6}$) (Table 3b). This argues strongly for linkage disequilibrium between the R141H allele and allele '16' of the marker.

Table 3 Linkage disequilibrium between the R141H mutation and allele '16' for marker D16S3020 in patients compound heterozygous for R141H. **A:** Allele frequencies of marker D16S3020 on R141H, and normal chromosomes in CDG-Ia patients; **B:** Linkage disequilibrium between the R141H mutation and allele '16' for marker D16S3020 in patients compound heterozygous for R141H

A

Allele	*R141H*	*Normal*
0	–	7
1	–	1
2	–	1
3	–	8
4	–	1
7	–	3
10	–	4
11	–	27
12	–	5
13	–	4
14	–	3
15	1	5
16	41	9
17	–	6
18	1	1
19	–	1
Total	43	86

B

	'16' present	*'16' absent*	*Total*
R141H chromosomes	41	2	43
Normal chromosomes	9	77	86
Total	50	79	129

Discussion

Our study clearly indicates a discrepancy between the frequency of the most prevalent *PMM2* mutation (R141H) and its occurrence in CDG-Ia.

On the basis of the observed carrier frequencies, between 1/14 400 (Danish group) and 1/25 000 (Dutch/Flemish group) homozygotes for R141H are expected under the Hardy-Weinberg equilibrium. Thus, one would expect to find the homozygous R141H/R141H genotype in 45% to 60% of the CDG-Ia patients. Not a single one has been found and this is statistically significant. The lack of homozygotes for R141H cannot be explained by genetic drift or non-random mating, but is easily explained by the severity of the mutation: the enzymatic activity of recombinant R141H protein is virtually zero.[8,13] Therefore, homozygosity for R141H is probably incompatible with life.

The occurrence of the disease is thus determined by the frequency of the other, rare mutations which is estimated to be between 1/300 and 1/400, depending on the population. The carrier frequency for the second most common mutation, F119L, which accounts for 30% (Dutch/Flemish) to 80% (Danish) of the non-R141H alleles must vary between 1/500 and 1/1000. The latter explains why we have not observed this mutation in our population study.

Given that homozygosity for R141H is probably lethal, the disease frequency is given by the frequency of compound heterozygotes for R141H and another mutation, and the occurrence of combinations of the other mutations. These calculated frequencies (1/18 000–1/20 000) are higher than the known frequency of the disease, which has been estimated at 1/40 000 in Denmark (unpublished observations) and 1/80 000 for other European countries.[14] Thus, we conclude that the disease is probably underdiagnosed.

Haplotype analysis indicated clear linkage disequilibrium between R141H and microsatellite marker D16S3020. Forty-one of 43 patients have the same marker allele, whilst this allele is present in only 10% of the normal population (Table 3a). The data strongly suggest that the patients share the same ancestral chromosome and that the mutation probably occurred only once, in the ancestral Caucasian population. The frequency of R141H in CDG-Ia patients is highest in US, Scandinavia, and in Western Europe, whilst it is less frequent among patients in Spain and Portugal. This trend is not yet significant. On the other hand, R141H has not been found in Japanese patients (Table 1).[4,7]

Because the CDG-Ia mutations are genetically lethal, they should slowly disappear unless the loss of alleles is compensated by new mutations or by a heterozygous advantage. The former is not the case because we have clearly shown that the R141H allele has the same genetic background in almost all the patients and the explanation must therefore be a heterozygous advantage for this mutation. This is reminiscent of the situation of the frequent ΔF508 mutation in cystic fibrosis (CF) patients. It has been proposed that CF heterozygotes may have a selective advantage against cholera-induced secretory diarrhoea[15] and/or resistance to typhus.[16] It is not clear yet what advantage a heterozygous state of a phosphomannomutase deficiency might be. In any case, the estimated effect is small (an advantage of approximately 0.6% is sufficient to maintain the frequency of the present level).

In summary, we report that one particular *PMM2* mutation, R141H, is far more frequent in the general population than expected on the basis of the frequency of the disease. Second, no homozygotes were observed and this severe mutation is probably not compatible with life in the homozygous state. The mutation shows a founder effect because of linkage disequilibrium, an observation which points to a small heterozygous advantage. The data allow us to give a rough estimate of the frequency of the disease: it could be as high as 1/20 000.

Acknowledgements

We are indebted to A Reuser and R Wevers for sharing DNA from a reference Dutch population and to G Van Camp and J Dumon for providing DNA from couples with miscarriages. We thank E Schaftingen and M Pirard for sharing data prior to publication (reference 8).

References

1 Martinsson T, Bjursell C, Stibler H *et al*: Linkage of a locus for carbohydrate-deficient glycoprotein syndrome type I (CDG1) to chromosome 16p, and linkage disequilibrium to microsatellite marker D16S406. *Hum Mol Genet* 1994; **3**: 2037–2042.
2 Matthijs G, Legius E, Schollen E *et al*: Evidence for genetic heterogeneity in the carbohydrate-deficient glycoprotein syndrome type I (CDG1). *Genomics* 1996; **35**: 597–599.
3 Matthijs G, Schollen E, Pardon E *et al*: Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). *Nat Genet* 1997; **16**: 88–92.
4 Matthijs G, Schollen E, Van Schaftingen E, Cassiman J-J, Jaeken J: Lack of homozygotes for the most frequent disease allele in carbohydrate-deficient glycoprotein syndrome type 1A. *Am J Hum Genet* 1998; **62**: 542–550.
5 Kjaergaard S, Skovby F, Schwartz M: Absence of homozygosity for predominant mutations in PMM2 in Danish patients with carbohydrate-deficient glycoprotein syndrome type 1. *Eur J Hum Genet* 1998; **6**: 331–336.
6 Bjursell C, Wahlstrom J, Berg K *et al*: Detailed mapping of the phosphomannomutase 2 (PMM2) gene and mutation detection enable improved analysis for Scandinavian CDG type I families. *Eur J Hum Genet* 1998; **6**: 603–611.
7 Kondo I, Mizugishi K, Yoneda Y *et al*: Missense mutations in phosphomannomutase 2 gene in two Japanese families with carbohydrate-deficient glycoprotein syndrome type 1. *Clin Genet* 1999; **55**: 50–54.
8 Pirard M, Matthijs G, Heykants L *et al*: Effect of mutations found in patients with carbohydrate-deficient glycoprotein syndrome Type IA on the activity of phosphomannomutase 2. *FEBS Letters* 1999; **452**: 319–322.

9 Van Tintelen P, Matthijs G, Braam W, Cassiman J-J, Duran M, Poll-The BT: Homozygosity for the F119L mutation in the PMM2 gene in an adult with CDG syndrome type Ia. *Eur J Hum Genet* 1998; **6** (Suppl 1): 62.
10 Dracopoli N, Haines J, Korf B *et al* (eds): *Current Protocols in Human Genetics*. John Wiley: 1997, pp 2.6.1–2.6.10.
11 Schollen E, Pardon E, Heykants L *et al*: Comparative analysis of the phosphomannomutase genes *PMM1*, *PMM2* and *PMM2*ψ the sequence variation in the processed pseudogene is a reflection of the mutations found in the functional gene. *Hum Mol Genet* 1998; **7**: 157–164.
12 Bjursell C, Stibler H, Wahlström H: Fine mapping of the gene for carbohydrate-deficient glycoprotein syndrome, type I (CDG1): linkage disequilibrium and founder effect in Scandinavian families. *Genomics* 1997; **39**: 247–253.
13 Kjaergaard S, Skovby F, Schwartz M: Carbohydrate-deficient glycoprotein syndrome type IA: Expression and characterisation of wild type and mutant *PMM2* in *E. coli*. *Eur J Hum Genet*, 1999; **7**: 884–888.
14 Kristiansson B, Stibler H, Hagberg B, Wahlström J: CDG a recently discovered hereditary metabolic disease. Multiple organ manifestations, incidence 1/80 000, difficult to treat. *Lakartidningen* 1998; **95**: 5742–5748.
15 Romeo G, Devoto M, Galietta LJ: Why is the cystic fibrosis gene so frequent? *Hum Genet* 1989; **84**: 1–5.
16 Prince A: The CFTR advantage-capitalizing on a quirk of fate. *Nat Med* 1998; **4**: 663–664.