Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing

Sharita Timal1,2, Alexander Hoischen3, Ludwig Lehle5, Maciej Adamowicz6, Karin Huijben2, Jolanta Sykut-Cegielska7, Justyna Paprocka8, Ewa Jamroz9, Francjan J. van Spronsen9, Christian Körner10, Christian Gilissen3, Richard J. Rodenburg2, Ilse Eidhof2, Lambert Van den Heuvel2, Christian Thiel10, Ron A. Wevers2, Eva Morava4, Joris Veltman3 and Dirk J. Lefeber1,2,

1Department of Neurology, 2Department of Laboratory Medicine, 3Department of Human Genetics and 4Department of Pediatrics, Institute for Genetic and Metabolic Disease, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands 5Department of Cell Biology and Plant Biochemistry, University of Regensburg, Regensburg, Germany, 6Department of Biochemistry and Experimental Medicine, The Children’s Memorial Health Institute, Warsaw, Poland, 7Department of Metabolic Diseases, Children’s Memorial Health Institute, Warsaw, Poland, 8Child Neurology Department, Medical University of Silesia, Katowice, Poland, 9Beatrix Children’s Hospital, University Medical Center of Groningen, University of Groningen, Groningen, The Netherlands and 10Center for Child and Adolescent Medicine and Center for Metabolic Diseases Heidelberg, Department Kinderheilkunde I, Heidelberg, Germany

Received February 10, 2012; Revised March 26, 2012; Accepted March 28, 2012

Congenital disorders of glycosylation type I (CDG-I) form a growing group of recessive neurometabolic diseases. Identification of disease genes is compromised by the enormous heterogeneity in clinical symptoms and the large number of potential genes involved. Until now, gene identification included the sequential application of biochemical methods in blood samples and fibroblasts. In genetically unsolved cases, homozygosity mapping has been applied in consanguineous families. Altogether, this time-consuming diagnostic strategy led to the identification of defects in 17 different CDG-I genes. Here, we applied whole-exome sequencing (WES) in combination with the knowledge of the protein N-glycosylation pathway for gene identification in our remaining group of six unsolved CDG-I patients from unrelated non-consanguineous families. Exome variants were prioritized based on a list of 76 potential CDG-I candidate genes, leading to the rapid identification of one known and two novel CDG-I gene defects. These included the first X-linked CDG-I due to a de novo mutation in ALG13, and compound heterozygous mutations in DPAGT1, together the first two steps in dolichol-PP-glycan assembly, and mutations in PGM1 in two cases, involved in nucleotide sugar biosynthesis. The pathogenicity of the mutations was confirmed by showing the deficient activity of the corresponding enzymes in patient fibroblasts. Combined with these results, the gene defect has been identified in 98% of our CDG-I patients. Our results implicate the potential of WES to unravel disease genes in the CDG-I in newly diagnosed singleton families.

INTRODUCTION

Genetic defects in the glycosylation of proteins and/or lipids result in a large and rapidly growing group of neurometabolic diseases, collectively called the congenital disorders of glycosylation (CDG). Of all glycosylation pathways, protein N-glycosylation is best understood (Fig. 1). The first part of the pathway, i.e. dolichol-glycan assembly and glycan transfer
to nascent proteins in the endoplasmic reticulum (ER), involves a multistep process and is identical for each N-glycosylated protein. Genetic defects in this process cause CDG type I (CDG-I). Patients often present with a highly variable multisystem phenotype, including neurological symptoms such as psychomotor retardation, muscle hypotonia, seizures, liver and kidney symptoms, endocrine and coagulation abnormalities and variable dysmorphic features (1,2). Due to their variability, clinical features are not discriminative for the many possible defects in the glycosylation route (3). Only in a few cases, genetic defects lead to specific clinical features, such as muscular dystrophy in DPM3-CDG (4), that can be used as a diagnostic criterion. The large number of possible genes involved in protein N-glycosylation has up to now hindered a direct sequencing approach for the identification of causative gene defects.

The current strategy for the identification of new disease-causing genes in CDG-I patients has heavily relied on yeast mutants and the functional knowledge of their involvement in the N-glycosylation pathway (5). A combined genetic-biochemical approach using homozygosity mapping in consanguineous families was successfully applied to identify novel genes in the N-glycosylation process (6,7). Altogether, these approaches have resulted in 17 genetically distinct CDG-I subtypes currently known (CDG-Ia to Iq, in the former nomenclature) that can be classified into four different functional groups (Fig. 1): dolichol-phosphate synthesis and recycling (A), synthesis and transport of nucleotide sugars (B), dolichol-linked oligosaccharide biosynthesis (C) and the oligosaccharyltransferase complex (D).

Whole-exome sequencing (WES) offers new opportunities to quickly identify disease genes in Mendelian disorders (8–11). Disease genes were successfully identified by prioritizing potential variants based on the data of homozygosity mapping and linkage analysis, or combining results from multiple patients with an identical clinical phenotype (12,13) or from patient–parent trios (14). The identification of disease genes in single cases remains difficult, although there has been some success for autosomal recessive disorders (15,16). Especially in single cases, the definitive annotation of a mutation as pathogenic is extremely challenging and requires the identification of independent cases with similar mutations or a functional confirmation. In this respect, glycosylation disorders have the advantage that functional assays can readily be chosen on the basis of the identified candidate gene.

In this paper, we applied WES on the remaining six unrelated, unsolved CDG-I patients from our cohort of 117 cases with different clinical phenotypes and from non-consanguineous parents. On the basis of the functional knowledge of the glycosylation process, a CDG-I gene list (Supplementary Material, Table S1) was used to prioritize the WES data of each individual CDG-I patient, which led to the identification of four of six defects. This shows the potential to apply WES for gene discovery in CDG-I patients.

**RESULTS**

**Patients and clinical presentation**

In our cohort of 117 CDG-I patients, six non-related patients without known consanguinuity remained unsolved. The patients represented a diverse clinical spectrum as briefly summarized in Table 1. Patient 3 was an adopted child of the Colombian
origin, while the other five patients were Caucasian. Four patients died in their infantile and childhood period, while two are still alive in their teens. Five patients presented with neurological symptoms, in three of them the clinical phenotype involved the central nervous system predominantly. No clear clinical symptoms could be identified that were directly suggestive of a known CDG-I subtype.

Biochemical and genetic diagnostics did not reveal a causative defect

Patient sera were screened by transferrin isoelectric focusing and showed abnormal N-glycosylation. For patients 1, 4 and 6, the clearly increased asialo- and disialotransferrin fractions suggested a CDG-I. For patient 2, the elevation of disialotransferrin in combination with a minor band of asialotransferrin and the absence of a protein polymorphism also showed a type I profile. For patient 3, a slight increase in monosialo- and trisialotransferrin was shown in addition to the pronounced increase in asialo- and disialotransferrin, whereas for patient 5, an additional band was observed for monosialotransferrin (Fig. 2). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of transferrin for all patients showed the loss of complete N-glycans, interpreted for all patients as CDG-I (data not shown).

Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown). Standard diagnostics included enzyme analysis of phosphomannomutase (PMM) and phosphomannose isomerase (PMI), followed by the analysis of lipid-linked oligosaccharides (LLO) by [3H]-mannose labeling of patient fibroblasts. All patients showed the production of Glc3Man9GlcNAc2-PP-dolichol, only for patient 1 in clearly decreased amounts. Analysis of the early LLO synthesis was performed by labeling of patient cells with [3H]-GlcNAc, interpreted for all patients as CDG-I (data not shown).
Table 2. Overview of all variants identified by exome sequencing in six CDG-I patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of variants</td>
<td>21,703</td>
<td>22,229</td>
<td>23,214</td>
<td>22,460</td>
<td>26,180</td>
<td>23,746</td>
</tr>
<tr>
<td>Quality filtered private variants of which:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop frame shift mutations</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Canonical splice site variants</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Highly conserved (PhyloP &gt; 3)</td>
<td>27</td>
<td>36</td>
<td>59</td>
<td>32</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>CDG-I candidate variants</td>
<td>0</td>
<td>1 (HMGCR:18%)</td>
<td>2 (PGM1:78%, ALG13:100%)</td>
<td>1 (ALG13:100%)</td>
<td>1 (PGM1:100%, TUSC3:41%)</td>
<td>2 (DPAGTI:63%, DPAGTI:15%)</td>
</tr>
<tr>
<td>CDG-I candidate genes according to recessive disease model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td>No clear hits</td>
<td>No clear hits</td>
<td>PGM1 with 78% variant reads</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To retrieve the quality-filtered private variants, variants from the single nucleotide polymorphism database and from our in-house variant database were excluded. Next, a quality filter was applied to remove all variants resulting in synonymous changes, or located in intergenic, intronic or UTR regions, or variants in less than five reads or with <15% variant reads.

Candidate genes in which variants are potentially homozygous or compound heterozygous (two variants per gene), defined as variants found in >80% for homozygosity or in 15–80% variants for heterozygosity.

Prioritization of exome sequencing data on the basis of a CDG-I gene list

DNA, isolated from patient fibroblasts, was used for WES using SureSelect human exome enrichment and sequencing on a SOLiD sequencer. The targeted exomes of the patients were on average covered by 37–51-fold. More than 76% of all targeted regions were covered at least 10-fold. Of the targeted CDG-I candidate gene regions, between 79 and 95% were covered at least 10-fold (Supplementary Material, Table S2). On average, 22,000 variants were identified per patient (Table 2). A list of potential candidates was created by excluding common variants as described and by excluding synonymous variants as well as all variants in intergenic, UTR and intronic regions. Next, quality filtering was applied by excluding variants found in less than five reads and variants detected in less than 15% variant reads, resulting in a list of 97–253 private variants. The remarkably larger amount of private variants for patient 3 may be explained by the fact that he is an adopted child from Colombian origin, an ethnic background that is inadequately represented in the single nucleotide polymorphism database (dbSNP), the 1000 Genomes project and our in-house database.

To further prioritize the remaining variants, we composed a list of potential CDG-I genes. Based on the four known subgroups, involved in CDG-I (Fig. 1), we added a broad group of additional genes not associated with CDG before (Supplementary Material, Table S1). Subsequent filtering of the private variants using this list resulted in eight variants in five patients (Table 2). Based on recessive inheritance, a subsequent prioritization was applied using the following criteria: a minimum of 80% variant reads for potential homozygous and hemizygous variants and between 15 and 80% for compound heterozygous variants. In this way, variants fitting with a recessive disease model were identified in three patients (Table 2), while confirmation by Sanger sequencing revealed an additional homozygous variant in patient 3. Homozygous, hemizygous and compound heterozygous variants were selected for segregation analysis in the family.

Identification of a novel CDG-I subtype due to a de novo mutation in X-linked ALG13

In patient 4, a missense mutation (c.280A > G; p.Lys94Glu) was found in the X-linked gene ALG13 in 100% of the reads, and hemizygosity was confirmed by Sanger sequencing. The mutation was not identified in DNA derived from the blood of the mother, while short tandem repeat marker analysis confirmed the maternal inheritance of the allele (data not shown), suggesting either a maternal germline mosaicism for the mutation or a de novo mutation in the patient. The mutated nucleotide was highly conserved (PhyloP value 4.32; Supplementary Material, Table S3). Lysine 94 is located in the C-terminal glycosyltransferase domain of ALG13 and is fully conserved down to yeast (Supplementary Material, Fig. S1). The corresponding amino acid in yeast (lysine 153) is positioned at the C-terminal end of the α5 helix which is thought to be involved in UDP-GlcNAc binding on the basis of structural comparison with the bacterial homolog MurG (17,18).

Deficiency of UDP-GlcNAc:GlcNAc-PP-dolichol GlcNAc-transferase activity in patient fibroblasts

ALG13 is the soluble subunit recruited by membrane-bound ALG14 to form the heterodimeric ALG13/ALG14 complex catalyzing the formation of GlcNAc2-PP-dolichol, the second step in the synthesis of the LLO precursor for the N-glycan assembly. The enzyme activity of UDP-GlcNAc:GlcNAc2-PP-dolichol GlcNAc-transferase was measured in fibroblasts by the incubation of a microsomal membrane preparation with non-radioactive UDP-GlcNAc as glycosyl donor and [14C]GlcNAc2-PP-dolichol as an acceptor. The elongation of [14C]GlcNAc2-PP-dolichol to...
GlcNAc2-PP-dolichol was found to be severely reduced when compared with control fibroblasts with a residual activity of 17% (Fig. 3B).

A novel CDG subtype due to mutations in PGM1

In patient 3, two heterozygous variants were identified using the CDG-I candidate gene list (Table 2). The variant in ALG14 (15.2% variant reads) was confirmed by Sanger sequencing to be heterozygous, while no other mutations could be found in the gene. Hence, this variant was excluded from further analysis. Sanger sequencing of the variant in PGM1 (c.415G>C, p.Gly121Arg, 78.4% variant reads) showed that this variant was in fact a homozygous mutation. Co-segregation in the family of the c.415G>c variant in PGM1 could not be determined because the patient was an adopted child. The nucleotide c.415G is highly conserved (PhyloP 3.25), and the mutation generates a stop codon resulting in a truncated protein missing the last 60 amino acids of the C-terminal domain (Supplementary Material, Table S3).

In patient 5, the variant in PGM1 (c.1507C>T; p.Arg503X) was found in 100% of the variant reads and was confirmed to be homozygous by Sanger sequencing and heterozygous in both parents and an unaffected sibling (Table 3). The nucleotide c.1507C is highly conserved (PhyloP 6.17), while the amino acid glycine 121 is fully conserved down to yeast (Supplementary Material, Table S3 and Fig. S2A).

Decreased phosphoglucomutase activity in fibroblasts from patients 3 and 5

To study the functional consequences of the mutations in PGM1, phosphoglucomutase activity was measured in patients' fibroblasts. Phosphoglucomutase converts glucose-1-phosphate to glucose-6-phosphate and is involved in the synthesis of cytoplasmic nucleotide sugars. Enzymatic activity was determined by assaying nicotinamide-adenine dinucleotide phosphate formation at A340. Enzyme activity values of patients and the lowest and highest enzyme activity values of 11 control fibroblasts are shown in the graph (Fig. 3E).
Mutations in DPAGT1 in patient 6 reveal a DPAGT1-CDG defect (CDG-Ij)

A compound heterozygous variant was identified in DPAGT1 (Table 2). The first variant (c.206T > A, p.Ile69Asn; 45.16% of the reads) was found in a highly conserved nucleotide (PhyloP 4.37) in exon 2. The affected amino acid isoleucine 69 is conserved during evolution down to tetraodon (pufferfish species), while the amino acid change is predicted to be damaging by Sift. This missense mutation was confirmed to be heterozygous in the patient and in the father and was absent in the mother and the unaffected brother (Table 3; Supplementary Material, Fig. S3A). The second variant (c.161 + 5G > A, 63.16% of the reads) replaced a highly conserved nucleotide (PhyloP 5.89) in the donor splice site of intron 1 (Supplementary Material, Table S3). The Sanger sequencing of genomic DNA confirmed heterozygosity in the patient and demonstrated the presence of the splice site variant in the heterozygous state in the mother and the unaffected twin brother. The variant was not detected in the father (Table 3; Supplementary Material, Fig. S3B). To further study the effect of the splice site mutation, mRNA was extracted from patient fibroblast cells, transformed into cDNA and sequenced by Sanger sequencing. The resulting cDNA sequence showed the c.206T > A variant in the homozygous form (Fig. 3G; Supplementary Material, Fig. S3C), suggesting the degradation of mRNA transcribed from the allele with the splice site mutation.

DPAGT1 encodes UDP-GlcNAc:dolichol-phosphate N-acetyl-glucosamine-1-phosphate transferase, catalyzing the transfer of GlcNAc-1-P from UDP-GlcNAc to dolichol-phosphate in the first step of LLO biosynthesis. DPAGT1 contains seven conserved domains, of which two are potential dolichol recognition sites (20). The missense mutation is located in the second amino acid of the highly conserved dolichol recognition motif F I/V X F/Y X X I P F X F/Y (21). Thus far, no other CDG-I gene defects have been identified with mutations in this dolichol-recognition motif.

Deficiency of UDP-GlcNAc:dolichol-phosphate N-acetyl-glucosamine-1-phosphate transferase activity in patient-derived fibroblast

In order to demonstrate DPAGT1 deficiency, a detergent solubilize from microsomal membranes was incubated with radioactive UDP-[14C]GlcNAc as a glycosyl donor and dolichol-P as an acceptor. As shown in Figure 3H, the formation of [14C]GlcNAc1-PP-dolichol was severely reduced compared with control fibroblasts with a residual activity of 22%.

No mutation detected in CDG candidate genes in two patients

For the remaining two patients (1 and 2), our strategy did not result in CDG-I candidate genes according to recessive inheritance. The CDG-I gene list was further applied directly to the unfiltered ~22 000 variants. Although additional heterozygous variants were found, they either were known as single

### Table 3. Overview of identified CDG-I defects

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gene name</th>
<th>Gene variant</th>
<th>Chromosome</th>
<th>Residual enzyme activities% to control</th>
<th>Identified mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>PGM</td>
<td>c.415G &gt; C</td>
<td>Chr1</td>
<td>−</td>
<td>Homozygous</td>
</tr>
<tr>
<td>4</td>
<td>ALG13</td>
<td>c.280A &gt; G</td>
<td>X</td>
<td>−</td>
<td>Homozygous</td>
</tr>
<tr>
<td>5</td>
<td>PGM</td>
<td>c.1507C &gt; T</td>
<td>Chr1</td>
<td>−</td>
<td>Homozygous/homozygous</td>
</tr>
<tr>
<td>6</td>
<td>DPAGT1</td>
<td>c.206T &gt; A</td>
<td>Chr11</td>
<td>−</td>
<td>Heterozygous/homozygous</td>
</tr>
</tbody>
</table>

Causative mutations as confirmed by Sanger sequencing, segregation analysis in available family members and residual enzyme activity measured in patient fibroblasts. Sequence analysis data are shown in Supplementary Material, Figures S1–S3. NA, not available; PGM, phosphoglucomutase. aSee Figure 3. Percentage of PGM enzyme activity is calculated as percentage of the mean control value. bGenomic position of the mutation based on the 2009 release of the UCSC genome browser (hg19). cThe splice site mutation led to wrongly spliced mRNA that most likely is degraded as indicated by the absence of its corresponding cDNA in Figure 3G.
nucleotide polymorphisms in public databases or were not in agreement with a recessive mode of inheritance.

DISCUSSION

In this study, WES was performed to solve the gene defect in a group of six patients with CDG-I. The application of a CDG-I gene list based on the functional knowledge of the N-glycosylation pathway led to the identification of the first ALG13-CDG (or CDG-Ir) due to an X-linked mutation, the first PGM1-CDG and compound heterozygous mutations in the postulated dolichol recognition domain of DPAGT1. These data demonstrate the potential of the list of CDG-I candidate genes to effectively select functionally relevant pathogenic mutations from WES data in singleton cases.

Defects in LLOs

Two of the identified defects (ALG13 and DPAGT1) catalyze the two initial steps of LLO biosynthesis, i.e. the incorporation of the first and the second GlcNAc residue, respectively. Assaying LLO synthesis via the [3H]-mannose labeling of fibroblasts or the measurement of cytosolic mannosyltransferase activities did not show any structural abnormalities. Indirect analysis of the GlcNAc transferase activities was performed by labeling with [3H]-GlcNAc, not showing a clear deficiency. These findings indicate that defects in the first part of dolichol-linked glycans assembly are difficult to diagnose using the current methods for short LLO analysis. Such assays are not yet well established and are only scarcely accessible, and the report of only a single case with a defect in these two initial steps of the pathway (DPAGT1-CDG) implies a limited experience with the interpretation of results. In addition, technical issues like the passage number of patient fibroblasts can influence the outcome (23). The direct assaying of enzyme activity in both patients was necessary to confirm a functional deficit. As described below, the clinical features of defects in LLO glycosylation are not discriminative and, together with the occurrence of multiple genes in this pathway necessitating a lengthy biochemical process, provides opportunities for the direct implementation of next generation sequencing (NGS) techniques.

Defects in nucleotide-sugar synthesis

Phosphoglucomutase is involved in the cytoplasmic biosynthesis of nucleotide sugars needed for glycan biosynthesis. The abnormal CDG screening results in patients 3 and 5 imply an influence on protein N-glycosylation. For the previously described single patient with a defect in PGM1, no glycosylation abnormalities have been reported (24). Another defect in nucleotide-sugar metabolism, classical galactosemia due to GALT mutations, also leads to abnormal protein N-glycosylation with a loss of complete N-glycans and truncated glycans, resulting in slightly different CDG-I screening profiles (25). As for galactosemia, the exact mechanism by which PGM1 deficiency leads to type I abnormalities is unclear. As will be described in more detail elsewhere, the minor bands of monosialo- and trisialotransferrin in patients 3 and 5 suggest the presence of incomplete glycans, in addition to the loss of complete glycans, as the characteristic of CDG-I. Therefore, the profile in these PGM1-CDG patients could best be described as CDG-I/II. The established enzyme assay provides a fast diagnostic method, in combination with specific clinical symptoms, to find novel patients with PGM1-CDG.

Clinical phenotyping

The elucidation of novel CDG-I gene defects allows a comparison of overlapping and distinctive clinical symptoms in our cohort and published cases. One interesting aspect was the presence of inverted nipples and fat pads in patient 1, which is often considered diagnostic in PMM2-CDG. The other interesting feature in some of our patients is dilated cardiomyopathy, an uncommon feature in CDG-I. The most helpful diagnostic symptoms were the presence of visual anomalies in association with central nervous system anomalies in our patients. To date, only a single case of DPAGT1-CDG was reported (22). The patient showed dysmorphic features, hypotonia and intractable seizures. Our patient shows in addition cataract, also reported in SRD5A3-CDG (6) and in ALG2-CDG (26). Patient 6 with ALG13-CDG shows bilateral optic nerve atrophy as was reported in SRD5A3-CDG and frequently observed in ALG1-CDG (27). In general, the short LLO biosynthesis defects show a more prominent neurological phenotype with severe epilepsy and ophthalmological involvement. Furthermore specific symptoms have been reported like increased occurrence of deafness in RFT1-CDG and ichthyosis in MPDU1-CDG, SRD5A3-CDG and DOLK-CDG. Furthermore non-syndromic dilated cardiomyopathy has recently been described as a clinical presentation of DOLK-CDG (28). For PGM1-CDG, the clinical phenotype of dilated cardiomyopathy and liver disease without neurological involvement should lead to the suspicion of this defect.

NGS for CDG-I gene discovery

Our results show the successful identification of four of six defects in singleton CDG-I families by a WES approach using the knowledge of the N-glycosylation pathway for variant prioritization. In the remaining two cases, this strategy failed to reveal a molecular diagnosis. One explanation for this is the incomplete detection of all coding variation by WES, due to incomplete exome enrichment kits, incomplete coverage, imperfect read mapping and variant calling. Alternatively, the incomplete knowledge of the genes influencing the glycosylation process may have hampered the correct prioritization of the disease-causing variants. An increasing number of studies show the impact of WES on disease gene identification. Confirmation of pathogenicity is most often done by a combination of bioinformatic predictions and comparison with clinically similar patients (29,30). Although functional assays are not readily available for many gene defects, a recent study on Wilson disease-related mutations in ATP7B showed that some of the published mutations did not lead to functional consequences (31). This stresses the importance of a functional confirmation of newly identified disease mutations, as we show in this paper.
With respect to the clinical and biochemical hurdles for CDG-I subtyping as described above, WES can directly be positioned in CDG-I gene discovery. Since PMI-CDG is a treatable disorder, the measurement of PMI should be initiated directly after the recognition of the presence of protein-losing enteropathy, especially in the absence of neurological symptoms. In view of the frequency of PMM2-CDG (Fig. 4A), PMM activity should be assayed after the exclusion of secondary causes for CDG-I. If PMM activity is normal, the clinical clues as described above (Fig. 4B) can be used to select accessible biochemical assays, followed by mutation analysis. If no further direct clinical clues are present, the analysis of LLOs by \[^{3}H\]-mannose labeling may be performed, which is relatively accessible in several laboratories in the world. Depending on availability, turn-around time and costs, WES could already be competitive at this stage of the diagnostic track (Fig. 4B). An approach using targeted enrichment arrays for known CDG-I genes prior to NGS was recently described (32). An advantage of WES or even better whole genome sequencing approaches is that they also allow the identification of novel CDG-I genes, as was highlighted in this study.

Further developments in NGS methodology will overcome potential technical pitfalls and, in combination with reduced costs and increased speed, are competitive for current diagnostics. In addition, it can be anticipated that additional CDG-I patients and novel defects will be identified via an exome-or genome-first approach resulting in an expanding clinical heterogeneity. In such cases, the prioritization of genomic variants using our CDG-I gene list will be helpful to identify the appropriate biochemical assay as required for the functional confirmation of the mutation.

In conclusion, the application of the current knowledge of the N-glycosylation pathway in the prioritization of WES data highly facilitated gene identification in CDG-I patients. This combined approach increased the amount of solved cases in our patient cohort from 95 to 98%, including the first X-linked ALG13-CDG and PGM1-CDG and a novel patient with DPAGT1-CDG. These results prompted us to propose a diagnostic workflow for CDG-I gene discovery including a prominent position for NGS techniques.

**PATIENTS AND METHODS**

**Patient description**

Blood and fibroblasts of patients were obtained for the diagnostics of inborn errors of metabolism and used after informed consent from parents and treating physicians. Clinical details of patients 1 and 6 have been described elsewhere as patients 2 and 9, respectively (33). The clinical symptoms of the six patients are presented in Table 1.

**CDG diagnostics**

Screening for CDG was performed by serum transferrin isofocusing as described (34). Secondary causes for CDG-I,
fructosemia, galactosemia and alcohol abuse, were ruled out in all patients. Biochemical assays for subtyping included enzyme assays for PMM, PMI and analysis of LLOs by the [3H]-mannose and [3H]-GlcNAc labeling of fibroblasts as described previously (35). In addition, enzyme assays for dolichol-phosphate-mannose synthase (DPM) and N-oligosaccharyltransferase were performed for patient 6.

Whole-exome sequencing
Genomic DNA was extracted from patient fibroblast pellets according to the manufacturer’s protocol using a Qiagen Mini kit (Qiagen) and was checked for DNA degradation on agarose gels. The exomes were captured and enriched using a SureSelect human exome enrichment kit (Agilent, Santa Clara, CA, USA; for patients 5 and 6, the 38-Mb kit was used, while for patients 1–4, the 50-Mb kit was used). After amplification, reads were sequenced on a SOLiD4 sequencer (Life Technologies, Foster City, CA, USA). Color space reads were iteratively mapped to the hg19 reference genome with the SOLiD BioScope software version 1.3. Small insertions and deletions were detected by the SOLiD Small indel tool. Called variants and indels were annotated using an in-house annotation pipeline (14,36) including hg19 genomic position, gene name, coverage number and percentage of variant reads, PhyloP nucleotide conservation score and Grantham score. Common variants were filtered out based on dbSNP31 and an in-house database of 300 exomes. Quality criteria were applied to filter out variants with less than five variant reads and less than 15% variation. Furthermore, synonymous variants, deep intronic, intergenic and UTR variants were excluded. The resulting list of private non-synonymous or splicing site variants was prioritized using a list, compiled of genes involved in the ER N-glycosylation pathway in the biosynthesis of dolichol-phosphate, dolichol-linked oligosaccharides, oligosaccharyltransferase complex, and in biosynthesis and transport of nucleotide sugars (Supplementary Material, Table S1), under the assumption of a recessive disorder.

Sanger sequence analysis
Primers flanking the candidate gene variants were designed with the Primer3plus program (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi, last accessed on 5 April 2012), were extended with an M13 tail to the 5′- and 3′-ends and were used for polymerase chain reaction (PCR). After the purification of the PCR product, the sequence reaction was performed with primers annealing to the M13 tails and using the bigDyeTerminator kit (Life Sciences, Krimpen aan de IJssel) according to the manufacturer’s instructions. The resulting fragments were analyzed on an Applied Biosystems ABI PRISM 3130xl Genetic Analyzer. Primer sequences are described in Supplementary Material, Table S4.

Enzyme assays
Skin fibroblasts were cultured in M199 medium (Life Technologies) supplemented with 10% fetal calf serum and 1% penicillin and streptomycin in a humidified atmosphere containing 5% CO₂ at 37°C. Cells were harvested with 0.25% Trypsin, washed three times with 0.9% NaCl and stored as pellets at −80°C before use.

Phosphoglucomutase activity
Fibroblast pellets (2 × 10⁶ cells) were resuspended in a 200 μl buffer consisting of 20 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) (pH 7.1), 25 mM KCl, 1 mM dithiothreitol (DTT), 10 μg/ml leupeptin and antipain and were homogenized on ice by three consecutive sonication steps of 8 s. The resulting homogenates were centrifuged (4°C, 8 min at 1550 g), after which the supernatant was stored for a maximum of 1 week at −80°C until the execution of enzymatic assays and protein determination. Phosphoglucomutase activity was measured as previously described (19), with minor changes. The reaction mixtures (150 μl of final volume) contained 50 mM HEPES (pH 7.1), 5 mM MgCl₂, 0.25 mM NADP, 0.5 mM glucose-1-phosphate, 1 mM glucose-1,6-bisphosphate, 10 μg/ml yeast glucose-6-phosphate dehydrogenase (Roche) and 5 μl of cell extract supernatant. The formation of NADPH was assayed spectrophotometrically (Konelab 20xTi) at A340 during 20 min of incubation at 37°C and control incubations without glucose-1-phosphate were subtracted. The protein level was determined with the U/CSF protein kit (Konelab™). Phosphoglucomutase activities were calculated as mU/mg protein.

UDP-GlcNAc:dolichol phosphate N-acetylglucosamine-1-phosphate transferase activity
Solubilized enzyme was obtained from a microsomal membrane fraction of fibroblasts, prepared as described (35). For solubilization, the protein concentration was adjusted to 6 mg/ml, 1.25 M sodium chloride, 0.5 mM dithiothreitol, 20% glycerol and 6 mM diheptanoyl-phosphatidylcholine, incubated for 20 min on ice and centrifuged for 40 min at 160,000 g. The reaction mixture for the determination of UDP-GlcNAc:dolichol phosphate N-acetylglucosamine-1-phosphate transferase (GPT) activity contained in a final volume of 0.06 ml: 28 mM Tris–HCl, pH 7.4, 19 mM MgCl₂, 0.7 mM sodium chloride, 0.7 mM DTT, 0.3% Nonidet P40 (NP-40), 3.5 mM diheptanoylphosphatidylcholine, 23% glycerol, 0.05 μCi UDP-[14C]GlcNAc (specific activity 317 mCi/mmol), 2 μg dolichol-P and solubilized enzyme (equivalent to 0.2 mg of membrane protein). After incubation for 3 min at 24°C, the reaction was stopped by addition of chloroform/methanol to give a ratio of chloroform/methanol/water of 3/2/1 v/v/v and processed by phase partitioning according to Sharma et al. (37). The glycolipid fraction was analyzed by thin layer chromatography on silica gel 60 plates (Merck) developed in chloroform/methanol/water (65:25:4 v/v/v). Radioactivity was detected and quantified by Phosphor-Imaging.

UDP-N-acetylglucosamine:GlcNAc1-PP-dolichol GlcNAc-transferase activity
A solubilized enzyme was obtained from a microsomal membrane fraction of fibroblasts, prepared as described (38). For solubilization, the protein concentration was adjusted to 6 mg/ml in 24 mM Tris–HCl, pH 7.5, 28% glycerol, 2.4 mM MgCl₂, 0.8 mM DTT, 1.5% NP-40 and incubated for 20 min on ice and centrifuged for 40 min at 160,000 g. The reaction mixture
for the determination of UDP-N-acetylglucosamine: GlcNAc1-P-dolichol GlcNAc-transferase (GNT) activity contained in a final volume of 0.06 ml: 27 mm Tris–HCl, pH 7.5, 27% glycerol, 10 mm MgCl₂, 0.8 mm DTT, 0.45% NP-40, 0.2 mm UDP-GlcNAc and [14C] GlcNAc1-PP-Dol (3000 cpm) and solubilized enzyme (equivalent to 0.1 mg membrane protein). After incubation at 24°C for the indicated time, the reaction was stopped by chloroform/methanol to give a ratio of chloroform/methanol/water of 3/2/1 (by volume) and processed as described above for GPT activity.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank the CDG patients and their families for their participation. We also thank Ir. Mailys Guillard for preparation of Figure 1 and Gerry Steenbergen-Spanjers for performing iso-electric focusing of transferrin.

Conflict of Interest statement. None declared.

FUNDING

This work was financially supported by the Institute of Genetic and Metabolic Disease (IGMD) (grant to D.J.L., R.J.R. and J.V.) and by the Ko¨rber-Stiftung to L.L.

REFERENCES


