The glucocerebrosidase locus in Gaucher’s disease: molecular analysis of a lysosomal enzyme

P K Mistry, T M Cox

Since the concept of lysosomal storage disorders was introduced by Hers in 1965 more than 30 inherited defects affecting this organelle have been recognised. Of these the most prevalent is the autosomal recessive condition Gaucher’s disease, which is caused by deficiency of glucocerebrosidase (EC 3.2.1.45). Rapid progress has been made in understanding the molecular basis of Gaucher’s disease and in the development of effective enzyme replacement therapy. Analysis of mutations that cause Gaucher’s disease will shed light on structure-function relationships of the glucocerebrosidase molecule and should define the influence of genotype on disease expression. Molecular analysis of the glucocerebrosidase gene can assist the genetic counselling of families affected by Gaucher’s disease and in the prognosis and treatment of individual patients with this condition.

Historical perspective
Gaucher’s disease was first described by Phillippe Gaucher in 1882, who reported a patient with hepatosplenomegaly in whom the characteristic abnormal macrophages were thought to represent a primary neoplasm of the spleen. Identification of the stored material as glucocerebrosidase in 1934 was followed many years later by the finding that glucocerebrosidase activity was markedly reduced in the spleen of patients with Gaucher’s disease by Brady et al and Patrick independently in 1965. Although the enzyme is deficient in most tissues of the body, it is the cellular specificity of the substrate that determines which tissues are principally affected. The substrate, glucocerebroside, is a constituent of complex glycolipids and is released in the breakdown of cell membranes, particularly leucocytes and erythrocytes, by mononuclear phagocytes. Thus, in Gaucher’s disease engorged macrophages accumulate throughout the body and cause hepatosplenomegaly, bone marrow infiltration with skeletal disease, and, rarely, involvement of the lungs and brain.

Classification of Gaucher’s disease
There are three main clinical subtypes of Gaucher’s disease. Most patients suffer from visceral involvement in the absence of neurological disease: this non-neuronopathic form of the condition is known as type I Gaucher’s disease. Within this group there is great variability in the rate of disease progression. Patients may present in childhood with hepatosplenomegaly, pancytopenia, bone necrosis, and osteoporosis or come to light in the ninth decade of life because of the incidental finding of splenomegaly. Type I Gaucher’s disease occurs rarely in all ethnic groups but it is more frequent in the Ashkenazi Jewish population. A wide range of disease activity is seen in the Jewish population but patients with late onset or mild disease are usually found to be of Ashkenazi descent. Type II (acute neuronopathic) disease is a rare disorder with no ethnic predilection. It causes a rapidly progressive neurovisceral storage disease and death in infancy. Affected infants come to light as a result of oculomotor abnormalities, cranial and bulbar nerve palsies, spasticity, convulsions, and opisthotonus. Type III (subacute neuronopathic) Gaucher’s disease progresses less rapidly but is associated with ataxia, myoclonus, and convulsions; there is progressive intellectual impairment which is accompanied by hepatosplenomegaly and skeletal disease. Death usually occurs in childhood or adolescence. This type of Gaucher’s disease is also rare but it is frequent in a genetic isolate that represents a single extended pedigree originating from the Norrbottian region of northern Sweden.

Diagnosis
Typically the diagnosis of Gaucher’s disease is established by the finding of characteristic glycolipid laden large macrophages on histological examination of bone marrow, the liver, or spleen. These findings are not absolutely specific since pseudo-Gaucher’s cells may appear in the marrow of patients suffering from a variety of conditions associated with increased cell turnover, including chronic myeloid leukaemia and thalassaemia. Determination of leucocyte acid β-glucosidase activity remains a reliable and simple means of confirming the diagnosis of Gaucher’s disease.

Deficiency of glucocerebrosidase results from defects in the enzyme protein itself but a
few patients have been reported in whom a deficiency of a heat stable activator (SAP2-sphingolipid activator protein 2) of glucocerebrosidase is responsible for a phenocopy of Gaucher's disease.11 Biochemical studies of the residual glucocerebrosidase activity in patients with Gaucher's disease, combined with the wide spectrum of clinical presentation, had led to the expectation that multiple mutant alleles would be responsible for the condition.12 Landmarks in the molecular analysis of Gaucher's disease include the purification of glucocerebrosidase from human placenta in 1973,13 the cloning of glucocerebrosidase cDNA by Sorge et al14 in 1985, and the complete sequence characterisation of the human glucocerebrosidase gene and its pseudogene by Horowitz et al15 in 1989.

**Glucocerebrosidase gene and pseudogene**

The structural gene for glucocerebrosidase and its pseudogene map to human chromosome 1q21.15 The structural gene contains 11 exons and spans 76 kilobases; the pseudogene is located 16 kb downstream. In the homologous regions, 96% nucleotide identity exists. Promoter regions of genes encoding lysosomal enzymes usually conform to the housekeeping model. The promoter of the glucocerebrosidase functional gene is unusual in this respect since it contains TATA and CAAT boxes but no Sp1 binding site.15 A promoter of the pseudogene can direct transcription of reporter genes and pseudogene transcripts have been isolated from extracts of normal cells maintained in culture.1617 These transcripts, however, cannot be translated since they contain multiple stop codons; in addition, exons are deleted during mRNA processing because of mutations in consensus splice sites. The pseudogene is smaller than the structural gene because of sequence loss in introns 2 (313 bp), 4 (626 bp), 6 (320 bp), and 7 (277 bp) as well as two missing exon sequences in exon 9 (55 bp) and exon 4 (5 bp). The missing intron sequences appear to be Alu sequences that have been inserted into the functional gene after the presumed duplication event that gave rise to the pseudogene.15 In addition, the pseudogene contains multiple point substitutions throughout its sequence: when transferred to the structural gene by gene conversion or recombination events, these mutations result in defects in glucocerebrosidase that cause Gaucher's disease. Such homologies and differences between the functional and pseudogene have important implications for investigation of mutations in Gaucher's disease and, as a result of transcription of the pseudogene, particularly affect the molecular analysis of glucocerebrosidase cDNA.

Complementary DNA encoding human glucocerebrosidase spans about 2 kb. The mRNA is unusual in that it has two functional AUG start codons (−57 and −118 nucleotides) but translation initiated at either site results in functional glucocerebrosidase.1418 Northern analysis indicates the presence of three glucocerebrosidase mRNA species of approximately 5-6, 2-5, and 2.0 kb. The longest mRNA species is thought to represent an incompletely spliced nuclear transcript whereas the shorter transcripts appear to arise from alternative sites for transcriptional initiation and polyadenylation.19 As expected for a lysosomal enzyme, a 19 amino acid hydrophobic signal sequence is present. The mRNA is translated into a protein that is cleaved to produce a mature polypeptide of 497 amino acids with a molecular weight of 55.5 kDa. The polypeptide contains five potential 0 linked glycosylation sites, four of which appear to be glycosylated. The active site of glucocerebrosidase resides in the C-terminal half of the molecule and investigations with covalently bound inhibitors suggest a critical role for asp283 and asp300 in catalysis.20

Numerous polymorphic sites have been identified in the glucocerebrosidase gene, eight in introns and three in the 5' untranslated region.21 These sequences are in linkage disequilibrium and constitute two main haplotypes. Various mechanisms have been proposed to explain the ascendency of these haplotypes. A search for further polymorphic sites might facilitate detection of carriers in relatives of patients in whom the nature of the mutation has not been established. Previous studies have identified a common polymorphism with respect to a PvuII restriction site at genomic nucleotide position 3931 in intron 6 which has been linked to many uncharacterised alleles in patients with Gaucher's disease.22

**Mutational analysis in Gaucher's disease**

**HETEROGENEITY OF MUTATIONS**

The multiplicity of mutations responsible for Gaucher's disease has long been suspected from the clinical heterogeneity of the condition and the different kinetic properties of the residual enzyme activity.12 Marked differences in the synthesis and processing of enzyme protein had also been shown,10 but the extent of the heterogeneity of lesions at the glucocerebrosidase locus has only recently become evident as a result of detailed molecular analysis. Excellent reviews, to which the reader is referred, catalogue these mutations.182124 In Ashkenazi Jews with type I disease, four mutations account for over 98% of disease alleles. These same four mutations account for approximately 60% of disease alleles in patients not known to have Jewish ancestry; rare or private mutations account for the remainder.2526 Many patients with Gaucher's disease exhibit compound heterozygosity for mutations at the glucocerebrosidase locus.

**COMMON MUTATIONS**

Molecular lesions in the glucocerebrosidase gene described hitherto include missense mutations, a splice site mutation, a nucleotide insertion, deletions, crossovers between the structural gene and pseudogene, and gene conversion events (non-homologous recombi-
The most prevalent mutation in non-
neuronopathic Gaucher’s disease is asp1448→ser
(N370S) which is caused by replacement of an
A by G at cDNA nucleotide position 1226.27
This mutation accounts for 75% of disease
alleles among Ashkenazi patients and approxi-
mately 25% among patients not known to be
Jewish.25 28 In vitro expression of this mutant
enzyme in insect cells indicates that, compared
with the wild type enzyme, it has reduced
specific activity and other abnormal catalytic
properties in relation to activation by nega-
tively charged phospholipids and SAP2.29
In Ashkenazi Jews the N370S mutation is linked
to a PoulI polymorphism; it is invariably
found in the context of the Pvl 1.1 genotype.22
The observation that most unknown alleles in
Ashkenazi patients occurred in the context of
the Pvl 1.1 genotype prompted Beutler et al50
to search for other widespread mutations in
this ethnic group. Recently they reported the
insertion of a single nucleotide, G, at cDNA
position 84, that results in a frameshift that
abolishes translation of glucocerebrosidase.30
This mutation has so far only been found as a
single copy in compound heterozygotes of
Ashkenazi descent and accounts for 13% of
disease alleles. A less common missense muta-
tion, leuc444→pro (L444P) is caused by replace-
ment of a T for a C at cDNA nucleotide 1448.31
This mutation occurs normally in the pseudo-
genotype and, thus, has the potential to complicate
molecular analysis of the structural gene in
Gaucher’s disease.14 The L444P mutation ac-
counts for only about 2% of disease alleles in
Ashkenazi patients but for approximately 40%
of the alleles in non-Jewish patients.32 34 The
mutation is associated with all three subtypes
of disease but tends to cause severe or neuron-
opathic disease when present in the homo-
yzous form. Leucine 444 is located in the
catalytic domain previously identified in the
enzyme by reaction with the substrate analogue
β-conduritol epoxide.20 Disruption of the
protein structure in this region produces an
unstable protein that possesses little residual
activity.33
Lately, Beutler et al44 have described a splic-
ing mutation (IVS2+1) that results in the
deletion of exon 2 from the mature transcript.
This accounts for approximately 3% of disease
mutations in unrelated Jewish patients.32
The IVS2+1 mutation (replacement of G by A at
genomic position 1067) is also normally pre-
sent in the glucocerebrosidase pseudogene.
Several mutant alleles of glucocerebrosidase
differ from the wild type sequence in as many
as four codons and correspond to sequence
variations found in the pseudogene. These
complex alleles represent chimeric molecules,
part functional gene and part pseudogene, that
may result from gene conversion events or
unequal crossing over.35 37
The most widespread missense mutations
responsible for Gaucher’s disease N370S,
L444P, and R463C (cDNA 1226G, 1448C and
1504T respectively) occur in residues at exons
9 and 10 which have been considered critical
for the formation of the active site. However,
other missense mutations occur throughout
the gene (four in exon 5, two in exon 6, one
in exon 7, five in exon 8, two in exon 11)38 39 40 and
our understanding of the part played by these
residues in the functional integrity of human
glucocerebrosidase will depend in the first
instance on the determination of the three
dimensional structure of the enzyme at atomic
resolution.

Mutational analysis in clinical practice
The human glucocerebrosidase gene has been
subject to intensive study in relation to
Gaucher’s disease. Given the prevalence of
Gaucher’s disease and the morbidity associ-
ated with its more severe forms, there is a need
for methods to facilitate prenatal screening and
diagnosis. The emergence of an effective al-
ternative to narrow transplantation in the
management of non-neuronopathic Gaucher’s
disease also has a bearing on diagnosis and
prognosis. Although enzymatic assay of acid β-
glucosidase activity may serve to confirm the
diagnosis of Gaucher’s disease when sus-
ppected, activities obtained in up to 20% of
obligate heterozygotes are within the normal
range of enzymatic activity. Under these cir-
cumstances, molecular analysis of the gluco-
cerebrosidase gene should aid the definitive
detection of carriers in families at risk for
Gaucher’s disease. In addition, procedures
based on the polymerase chain reaction for the
analysis of genomic DNA should improve the
potential for early prenatal diagnosis by chor-
ionic villus sampling where there has been
experience of neuronopathic Gaucher’s disease
(figure).
Molecular analysis of the glucocerebrosidase
gene is complicated by the presence of the
transcribed pseudogene and thus methods to
detect mutations must either amplify the
structural gene sequences selectively or allow
separation of the amplified structural gene
from the amplified pseudogene sequences.28 30
Screening for known mutations has relied on
hybridisation to allele specific oligonucleotide
probes or, more simply, when the mutation
alters a restriction site, by digestion with ap-
propriate restriction endonucleases. Examples
of the latter include the NciI site associated
with the L444P mutation,31 the HhaI site asso-
ciated with the P415N mutation,39 and deletion
of a CfrI 10 site in association with N463C.25
However, most of the mutations that cause
Gaucher’s disease do not alter restriction sites
and screening for these has been facilitated by
the application of the ‘mismatch PCR’ tech-
nique.40 The search for known mutations,
regardless of whether or not a restriction site is
present, has been further simplified by the use
of the amplification refractory mutation sys-
tem (ARMS) which permits rapid genotyping
based on allele specific amplification in the
PCR in the single step45 (figure). These pro-
cedures allow genetic diagnosis of Gaucher’s
disease to be carried out in diagnostic labora-
tories outside major research centres. In our
laboratory, using seven Ashkenazi Jewish
patients with Gaucher’s disease, 11 disease
alleles were identified as N370S, and two as
84GG, leaving only one unknown allele. Among our other 17 patients from other ethnic backgrounds, ARMS analysis for common mutations showed six N370S alleles, nine L444P alleles, two N463C alleles, and one IVS2+1 leaving less than half the disease alleles in this diverse group of patients unidentified.

**Population genetics**

Gaucher’s disease appears to be most common in Ashkenazi Jews in whom a recent survey has suggested that the frequency of homozygosity for mutations may be as high as 1 in 976.1 This is largely because of the prevalence of the N370S mutation which accounts for 75% of disease alleles.28 Other common mutations in this population are 84GG (13%), IVS2+1 (3%), leaving less than 5% of alleles unknown.294 As a result of studies in more than 2000 Ashkenazi subjects, Beutler et al41 estimated the gene frequency for N370S and the 84GG mutation to be 0.032 and 0.00217, respectively; thus 1 in 980 offspring of unselected marriages between Ashkenazi Jews are predicted to be N370S homozygotes. In their survey, Beutler et al discovered four asymptomatic persons harbouring two copies of the N370S mutation when they examined a random Ashkenazi sample. The N370S mutation was found to be more prevalent than expected in the healthy Ashkenazi population in relation to a large sample of Ashkenazi patients with Gaucher’s disease. Beutler et al estimated that approximately two-thirds of the alleles were missing from the patient population with Gaucher’s disease, once again indicating variable expressivity of the homozygous form of this mutation and its association with relatively mild disease. This mutation exists at polymorphic frequency among Ashkenazi Jews and raises the possibility that it confers a selective advantage in the heterozygous state. The N370S and 84GG mutations are found consistently in the context of their respective haplotype associations, an observation that suggests that they arose on a single ancestral chromosome and were selected for or spread by diffusion from a single founding population. The Swedish isolate found in Norrbotten is strongly associated with type III Gaucher’s disease.5 All 10 patients studied were found to be homozygous for L444P and careful pedigree analysis later showed that these subjects were descended from a common ancestor in northern Sweden.42 Among non-Jewish patients, L444P is the most common disease allele and accounts for approximately 40% of disease alleles. In the general population, L444P occurs in the context of different haplotypes, a finding that suggests that this pathological variant has arisen repeatedly as a result of distinct mutational events.23

**Genotype-phenotype correlations in Gaucher’s disease**

Many inherited diseases show variable clinical expressivity and Gaucher’s disease is no exception: type II (neuronopathic disease) is a condition of infancy that progresses rapidly to death whereas other persons homozygous for other mutations in the glucocerebrosidase gene may remain asymptomatic in their ninth decade of life. Striking heterogeneity of clinical expression may also be observed in full sibs within a given pedigree affected by Gaucher’s disease and provides clear evidence that factors outside the glucocerebrosidase locus may also influence the degree to which the enzymatic deficiency is responsible for clinical manifesta-
The glucocerebrosidase locus in Gaucher's disease: molecular analysis of a lysosomal enzyme

893

The glucocerebrosidase
Gaucher's disease: assignment
median
Among those
ence or and,
of asymptomatic
activity may indeed
influence the severity of the disease but which are not yet understood. The role of
variation in the expression or quality of SAP 2
in relation to disease expression has yet to be systematically investigated.

Homozygosity for the L444P mutation causes severe Gaucher's disease and correlates
with neurological complications. Here again, however, striking variability in phenotype has
been observed ranging from fulminant neurological involvement characteristic of type II
disease to the more subacute neuronopathic form in type III. A few patients with the type I
phenotype have been found to be homozygous for the L444P mutation.45 In some type II
patients who were originally designated L444P homozygotes, numerous additional mutations
in the glucocerebrosidase gene resulting from crossover events with the pseudogene have
been found.46-47 The correlation with neurological complications of Gaucher's disease
does not appear to apply in all population groups, however. In Japanese patients homozygosity
for L444P has been found among the non-neuronopathic forms of Gaucher's disease
and a different mutation has primarily been correlated with type III disease (pch21-3 ile)
(F213I).46,47 Recent investigations of the effect of the R463C mutation illustrate another
aspect of genotype-phenotype correlations in Gaucher's disease. In two studies R463C has
been found to be associated with severe disease and neurological involvement,48,49 but in
another investigation it is associated with mild disease,50 yet when this mutant enzyme
is expressed in insect cells, its specific activity has been found to be little impaired. However,
the enzyme is also much less susceptible to activation by the sphingolipid activator protein
in vitro.51 Thus, in any individual patient it is difficult to predict accurately the outcome of
disease solely from knowledge of the genotype. To determine the origin of phenotypic variation is
a scientific question of importance which may eventually lead to therapeutic approaches to
disease modification, for example, in relation to the sphingolipid activator protein. The
investigation of the source of phenotypic variation will be greatly helped by the development
of methods to measure residual glucocerebrosidase activity in situ and this might be
achieved by loading lysosomes in macrophages obtained from patients with Gaucher's disease
with radiolabelled glucocerebroside or a suitable fluorescent artificial substrate.52 This
would in the first place allow the hypothesis of Conzelman and Sandhoff53 to be examined. In
the same way, the production of models of Gaucher's disease in experimental animals54 may
also assist in the investigation of the cellular pathophysiology of glucocerebrosidase
deficiency in man.

We wish to thank Drs Ian Ellis and Anthony Fensom of the Paediatric Research Unit,
UMDS-Guy's Campus, London SE1 9RT for referring the family depicted in the figure.

In general, it may be said that the more severe the disease the lower the level of resi-
dual enzymatic activity that can be expected. The level of activity needed to protect
against the more severe manifestations of Gaucher's disease may be a small fraction of
the normal. Conzelman and Sandhoff50 proposed that there exists a critical threshold in
the activity of lysosomal enzymes below which accumulation of undegraded substrate occurs.
This critical threshold and the rate of accumula-
tion of undegraded substrate may differ between cells, that is, between Kupffer cells
and splenic macrophages or between the same cells at different developmental stages in
different tissues, depending on the rate of influx of substrate. Thus a very small difference of
residual activity could markedly affect the age of onset of the disease, the rate of its progres-
sion, and its clinical manifestations. An additional factor in relation to Gaucher's disease is
that, during its course, when the spleen reaches a critical size, hypersplenism may re-
sult in the dramatic increase in glycolipid turnover from blood cell membranes. This
would be expected to accelerate the deposition of undegraded substrate. Thus secondary
pathophysiological disturbances may influence the apparent severity of Gaucher's disease
disproportionately and lead to bias in the cor-
relation of given genotypes with clinical mani-
festations. Other mechanisms may influence
the activity of glucocerebrosidase in Gaucher's
cells: there may be enhanced transcription,
resulting in a 2 to 3 fold increase in steady state
mRNA that would partially compensate for
decreased enzymatic activity52 and partial amelioration of the disease.

Because of the implications for prognosis and counselling of carriers, an important ob-
jective in the molecular analysis of the glucocerebrosidase gene has been to correlate
the nature of the mutation with the clinical features of the disease. As indicated above,
precise correlation has been difficult to achieve in Gaucher's disease. Nevertheless, it is clear
that genotype does influence phenotype strongly. Indices of disease severity, that is,
age of presentation or a composite score based on other clinical criteria, have been used to
estimate rates of disease progression.44 In general, homozygosity for the N370S mutation
confers the mildest form of disease and, al-
though the presence of another allele is usually associated with more aggressive features,
expression of one copy of this mutated gene may indeed provide sufficient enzymatic ac-
tivity to preclude development of neurological disease. Some homozygotes for N370S remain
asymptomatic and have very few manifestations of the disease even on careful clinical
study and, as indicated above, up to two-thirds of persons with this genotype may escape
medical attention for Gaucher's disease.44 Among those who present with symptoms, the
median age of detection is 30 years but a few subjects have a disease with onset in adoles-
cence or early adult life. In one such case the presence of a deleted allele resulting in incor-
rect assignment of a homozygous genotype was
later shown.44 Despite these exceptions there are clearly genetic and environmental factors
that influence the severity of the disease but which are not yet understood. The role of
variation in the expression or quality of SAP 2
in relation to disease expression has yet to be systematically investigated.
for prenatal diagnosis and the Wellcome Trust for support. Mrs Joan Grantham kindly rendered the manuscript suitable for publication.