Mucopolysaccharidosis III: Molecular genetics and genotype-phenotype correlations

JG Gilkes¹, BD Patterson¹, CD Heldermon¹

Abstract

Introduction

Sanfilippo Syndrome or Mucopolysaccharidosis III (MPS III) is a group of lysosomal storage diseases resulting from a deficiency of one of four lysosomal enzymes: Type A - heparan N-sulfatase (SGSH), Type B - α-N-acetylgalactosaminidase (NAGLU), Type C - acetyl CoA α-glucosaminide acetyltransferase (HGSNAT) and Type D - N-acetylgalactosamine-6-sulfatase (GNS). Each of these enzymes is necessary for degradation of heparan sulphate. Deficiency of any of these enzymes manifests as a neurodegenerative disorder with accompanying somatic manifestations. Currently treatment is limited to supportive care. MPS IIIA and IIIB are the most common subtypes of MPS III and will be further discussed in this review. The integral genes underlying both these diseases have been cloned and characterized. Through genetic analysis of the cDNA from MPS IIIA and B, researchers have begun to link many genetic mutations to their resultant phenotypes, and discern geographic differences in mutational variation. Here, we highlight many of the known MPS IIIA and B mutations and present them in the context of ethnic and geographic differences in an attempt to discern genotype-phenotype correlations and patterns of inheritance.

Conclusion

Most mutation sites have variable severity. A few sites have predictably more or less acute disease courses but all described mutations still result in progressive neurodegeneration and premature death.

¹ University of Florida, Gainesville, FL, USA

Introduction

Mucopolysaccharidoses are a group of eleven inherited lysosomal storage diseases (LSDs) resulting from a particular enzyme deficiency. Sanfilippo Syndrome, commonly referred to as Mucopolysaccharidosis III (MPS III), results from a deficiency in one of four lysosomal enzymes needed to break down heparan sulfate (HS). Heparan sulphate functions biologically as a proteoglycan which occurs as cell-surface and extracellular matrix macromolecules. These proteoglycans play crucial roles in regulating key developmental signalling pathways by binding to specific protein ligands and in maintaining cellular homeostasis. MPS III is inherited in an autosomal recessive manner and each of the four enzyme deficiencies defines a particular type of MPS III. Namely, Type A - heparan N-sulphatase (SGSH, OMIM # 252900), Type B - α-N-acetylgalactosaminidase (NAGLU, OMIM # 2529520), Type C - acetyl CoA α-glucosaminide acetyltransferase (HGSNAT, OMIM # 252930) and Type D - N-acetylgalactosamine-6-sulfatase (GNS, OMIM # 252940). Deficiency of any of these enzymes manifests as a neurodegenerative disorder with accompanying somatic manifestations. Of the mucopolysaccharidoses, MPS III is the most frequent, with an estimated incidence of 0.28 - 4.1 per 100,000 persons¹. MPS IIIA and -B, are the most common subtypes of MPS III²,³ and will be further discussed in this review.

The cDNA sequences for SGSH⁴ and NAGLU⁵,⁶ have been cloned and characterized, with gene locations on chromosomes 17q25.3 and 17q21.2, respectively. The SGSH sequence contains eight exons which span approximately 11 kb and encodes a 502 amino acid protein with five potential N-glycosylation sites located at positions 41, 142, 151, 264 and 413 (neXtProt, PX51688). Whereas, the NAGLU sequence contains six exons which span approximately 8.3 kb and encodes a 743 amino acid protein consisting of a 20- to 23- residue sequence considered to be the signal peptide and six potential N-glycosylation sites located at positions 261, 272, 435, 503, 526 and 532 (neXtProt, PX54802).

Once synthesized, these enzymes are shuttled to the trans-golgi network where mannose-6-phosphate (M6P) modifications are added and serve as ligands to interact with M6P receptors. This interaction allows lysosomal enzymes to be segregated from other proteins and transported to lysosomes under normal conditions. The aim of this review is to aggregate and synthesize retrospective genetic analysis data on MPS IIIA and -IIIB with the goal of highlighting genotype-phenotype correlations and clinical predictive patterns associated with disease severity.

Discussion

Natural History

The clinical evolution of Sanfilippo syndrome is typically divided into three stages. Following an initial period of normal development, the first phase of the disease is characterized by a delay in cognitive development, most notably speech, and becomes apparent between two and six years. During this phase parents often report delays in development of walking and talking, increased infections, hernias and diarrhoea⁷. The second phase of the disease is characterized by intellectual decline and regression of developmental milestones with increased behavioural problems which may become apparent as early as three to four years of age. At diagnosis, speech development is generally much more delayed than motor development.
Coarsening of the face with increased hair coarseness and development of organomegaly may become more apparent\(^1,3,6,8,9,10\). During the third and final stage behavioral problems slowly resolve, accompanied by the onset of severe dementia and decline of all motor functions, leaving patients in a vegetative-like state. Death, usually due to respiratory tract infections, generally follows at the end of the second to third decade of life\(^2,11\).

There are, however, some reports of survival into the late sixties in those with an attenuated phenotype\(^12,13\).

Extensive heterogeneity exists within the MPS III population resulting in diverse phenotypic presentations, thus making definitive clinical predictions of disease course difficult. Phenotypes are often characterized as being attenuated, intermediate or severe based on clinical course. In general, MPS IIIA is considered to be more severe than MPS IIIB\(^14,15\). On the other hand, MPS III B shows more variation in clinical outcome\(^3,16\), even among members of the same sibship\(^15\).

**Molecular Genetics**

**MPS IIIA**

The three dimensional structure of SGSH has not been fully elucidated. Over 120 mutations have been reported in the SGSH gene, the majority of which are missense mutations. These mutations are scattered among the eight exons of the SGSH protein (Figure 1).

Exon locations are as follows: E1, 1-30 aa; E2, 30-83 aa; E3, 84-119 aa; E4, 119-169 aa; E5, 169-221 aa and E6, 222-249 aa; E7, 249-317 aa; and E8, 317-502 aa. The ability to predict clinical disease course based on mutational analysis of predictive markers represents a gold standard.

Many studies have begun to map mutations to particular locations on the SGSH protein. A meta-analysis of such studies has highlighted some genotype-phenotype correlations. While clear predictive trends are difficult to discern based on current data, it is noteworthy that several missense mutations associated with the severe phenotype have been mapped to E2 (Figure 1). Further, many mutation locations are associated with multiple amino acid substitutions which may result in modified disease severity (Table 1). These preliminary observations may hold future utility in clinical disease prediction.

Further, numerous nonsense mutations, insertions, deletions and polymorphisms have been reported in the SGSH gene (data not shown).

**MPS IIIB**

Over 150 mutations have been reported in the NAGLU gene to date. With the increased use of genetic sequencing technology, it is likely that the number of associated mutations will continue to increase. Studies aimed at elucidating genotype-phenotype correlation patterns may influence clinical predictive capabilities. However, to date, few such correlations have been made. This is attributed to the considerable genetic heterogeneity which has been described in MPS IIIB patients of different origins. The majority of these mutations are missense mutations which are scattered among each of the six NAGLU exons (Figure 2).

Exon locations are as follows: E1, 1-128 aa; E2, 128-177 aa; E3, 178-226 aa; E4, 227-255 aa; E5, 256-341 aa and E6, 341-743 aa. The N-terminal domain of the NAGLU protein, containing the signal sequence, is located within E1, while the central domain, 130-467 aa, is found within E2-E6. The C-terminal domain is located within E6, 473-734 aa (NetProt.org). While no clear predictive disease severity trend emerges, it is of note that many mutations associated with a severe phenotype are located on exon six, with a small cluster emerging between amino acid sequences 501 and 521. Interestingly, several mutations were reported to occur at CpG islands within exon six, which are known to be mutational hotspots. Zhao and colleagues identified five mutations occurring at CpG islands, one deletion and four missense, using SSCP DEFINE analysis of PCR-amplified segments of genomic DNA from patients with Sanfilippo syndrome B. These mutations, R297X, R626X, R643H and R674H, resulted in the replacement of arginine to a stop codon or histidine\(^5\).

Further, several nonsense mutations, insertions, deletions and polymorphism events have been reported in this exon.

---

**Figure 1:** The majority of mutations in the SGSH gene are missense mutations and occur at numerous amino acid (aa) positions in each exon (green). Several aa residues appear to be hotspots for mutations and are associated with several aa substitutions (*, Table 1). Some genotype-phenotype correlations have been proposed, severe (red), intermediate (blue) and attenuate (pink); however, most missense mutations have an uncharacterized phenotype (black).

**Critical review**
Gilkes JA, Patterson BD, Heldermon CD. Mucopolysaccharidosis III: Molecular genetics and genotype-phenotype correlations.


Figure 2: The majority of mutations in the NAGLU gene are missense mutations and occurs at numerous amino acid (aa) positions in each exon (green). Mutational hotspots are indicated by (*, Table 1). Some genotype-phenotype correlations have been proposed, severe (red), intermediate (blue) and attenuated (pink); however, most missense mutations have an uncharacterized phenotype (black).

Genotype-Phenotype Correlations and the Interplay of Ethnicity

Currently, over 120 mutations each in the SGSH and NAGLU genes have been reported and added to the Human Gene Mutation Database (HGMD, http: //www.hgmd.org). The majority of these mutations are missense mutations while others include nonsense, deletions, insertions and splice site mutations.

In general there is poor correlation between genotype and clinical phenotype. This may partly be attributed to low allele frequencies among the different mutations. Studies suggest that heterogeneity in disease severity as a consequence of different genetic mutations contributes to the wide spectrum of clinical phenotypes. Further, the majority of mutations are unique to a single individual or individual family. Predictive correlations may also be hampered by numerous polymorphisms that may modify disease severity.

Nevertheless, several researchers have begun to tease apart such correlations. Weber and colleagues identified a common mutation in the SGSH gene in forty five unrelated MPS IIIA patients from the Netherlands. The R245H was present in fifty one alleles which represented over fifty six percent of the total allelic population.

This pathogenic mutation was also found to be prevalent in the Polish population and was associated with the severe phenotype. Similarly, the S66W mutation was also found to be prevalent in this population and was also associated with the severe phenotype. Interestingly, patients compound heterozygous for the S298P mutation in combination with one of the mutations associated with the aforementioned classical severe phenotype had a significantly longer preservation of psychomotor functions and a longer survival.

Furthermore, in a study of fifteen British participants, the R245H mutation had a correspondingly high frequency of twenty percent. However, the majority of these mutations were particular to an individual family. Noteworthy, R245H, combined with R74C was also found to be the prevalent mutation in German and Polish populations, with a frequency of greater than fifty percent.

In a subsequent study of twenty four Italian MPS IIIA patients, the S66W mutation occurred at the highest frequency of thirty three percent and was shared among six patients originating from the same region, thereby suggestive of a common founder.

Subsequently, Beesley and colleagues assessed twenty three patients from the United Kingdom and identified six mutations found in more than one unrelated persons, S66W, R74C, R245H, 1091delC, 1156ins6, and V486F. The R74C, R245H, S66W, and 1091delC are known to be prevalent in Polish, Dutch, Italian, and Spanish populations, respectively. Together, these six mutations accounted for more than sixty six percent of the mutant alleles in this study. The large, shared presence of the R245H and S66W mutations in these European regions may be suggestive of a founder’s effect. It is possible that these mutations may represent a screening target and predictive tool for disease severity in this population. It is also of note that several polymorphisms associated with an unaffacted phenotype have been identified in the SGSH gene, including R456H, which has a high frequency of fifty five percent in the general Australian population.

Licensee OAPL (UK) 2014. Creative Commons Attribution License (CC-BY)

patients with MPS IIIA four missense mutations and five polymorphisms were identified. All five polymorphisms were in Hardy–Weinberg equilibrium, suggesting little or no recombination in the SGSH gene. By combining this information with ethnic background of patients, new screening practices may be on the horizon for MPS IIIA patients.

Evidence suggests that there is more clinical variation in MPS IIIB patients, therefore identifying a genotype-phenotype relationship may prove difficult. To investigate this potential relationship, Weber and colleagues conducted a retrospective study of 40 Dutch MPS IIIB patients, and noted that only five mutations were found in more than one patient. R643C and R297X each accounted for around twenty percent of MPS IIIB alleles in the Dutch patient group, while P521L, R565W and R626X each had a frequency of about six percent in Australasian patients. R643C seemed to be a uniquely Dutch allele and clearly conferred the attenuated phenotype. Weber also noted that several arginine residues seem to be 'hot-spots' for mutations.

Mutational analysis of seven Japanese families revealed that the two participants exhibiting the most severe phenotype were homozygous for R482W, and R565P, respectively. Participants exhibiting an attenuated phenotype were compound heterozygous for F314L and R565P. Interestingly, the homozygous R565W mutation was shown to be associated with an intermediate form of the disease, thereby highlighting unique role of mutation variation in the modulation of disease severity. Tanaka further suggested that the R565P mutation is common in Okinawa.

This was later corroborated by Chinnen and colleagues who showed that five participants from Okinawa also bore the R565P mutation, thereby suggesting a founder effect. Subsequently, in a cohort of eleven MPS IIIB Portuguese patients, the R234C mutation attained the highest prevalence of thirty two percent of mutated alleles. This particular mutation is also common in Spanish populations, thereby suggesting a common origin. In a recent study of 136 consanguineous families (90% Iranian, less than 10% Turkish or Arabic), three of four children, born to parents related as first cousins once removed exhibited the severe form of MPS IIIB and were homozygous for the R565Q missense mutation. Strikingly, in these varying ethnic and geographic populations, mutation of R565Q primarily resulted in the severe phenotype of the disease. Lastly, mutation analysis was conducted on twenty one severely affected Greek MPS IIIB patients from eighteen different families. In this population, Y140C, H414R, and R626X account for approximately seventy percent of the studied alleles. Elucidating the ethnic and geographical similarities in amino acid mutation allelic frequencies may hold the key to understanding the genetic history of individual patients, and may improve carrier detection and genetic counselling in affected families.

Conclusion

The vast majority of affected individuals with MPS IIIA&B carry a unique mutation but neither prediction of certain mutations occur. Generally elusive. As next generation sequencing techniques becomes less expensive and more widely used, undoubtedly more mutations will be identified and perhaps poly-morphisms will enable identification of which sites are more and less severe and eventually allow better identification of the population prevalence of these diseases.

More importantly as enzyme replacement, substrate reduction methods, and gene therapy approaches move from the bench to the clinic, identification of these diseases early will allow earlier and more effective treatment.

References
4. Scott HS, Blanch L, Gou XH, Freeman C, Orsborn A, Baker E, Sutherland GR, Morris CP & Hopwood JJ. Cloning of the sulphamidase gene and identification of

Table 1: Several amino acid substitutions occur in the same mutation location, both in the SGSH (A) and NAGLU (B) proteins (nexTProt, NX.P51688 and NX.P54802, respectively). These substitutions may result in varying

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Amino Acid</th>
<th>Amino Acid</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Substitution</td>
<td>Position</td>
<td>Substitution</td>
</tr>
<tr>
<td>D32</td>
<td>E, G</td>
<td>F48</td>
<td>C, L</td>
</tr>
<tr>
<td>R74</td>
<td>C, H</td>
<td>G79</td>
<td>C, S</td>
</tr>
<tr>
<td>R150</td>
<td>W, Q</td>
<td>R130</td>
<td>C, H</td>
</tr>
<tr>
<td>R206</td>
<td>P, H</td>
<td>Y309</td>
<td>C, H</td>
</tr>
<tr>
<td>V220</td>
<td>M, L, A</td>
<td>V334</td>
<td>F, I</td>
</tr>
<tr>
<td>D235</td>
<td>V, N</td>
<td>R482</td>
<td>W, Q</td>
</tr>
<tr>
<td>F288</td>
<td>S, L</td>
<td>R565</td>
<td>L, Q, W, P</td>
</tr>
<tr>
<td>P293</td>
<td>T, S</td>
<td>R582</td>
<td>T, P</td>
</tr>
<tr>
<td>Q307</td>
<td>P, E</td>
<td>R643</td>
<td>C, H</td>
</tr>
<tr>
<td>R377</td>
<td>C, H</td>
<td>R674</td>
<td>C, H</td>
</tr>
<tr>
<td>R443</td>
<td>P, W</td>
<td>R737</td>
<td>G, S</td>
</tr>
</tbody>
</table>


