Available online at www.sciencedirect.com
ScienceDirect
Molecular Genetics and Metabolism 89 (2006) 139–149
Molecular Genetics and Metabolism
www.elsevier.com/locate/ymgme

Determinant factors of spectrum of missense variants in mucopolysaccharidosis IVA gene

Shunji Tomatsu [a,*,1], Adriana M. Montaño [a,1], Patricia Lopez [a], Georgeta Trandafirescu [a], Monica A. Gutierrez [a], Hirotaka Oikawa [a], Tatsuo Nishioka [a], Matheus B. Vieira [a], Tadao Orii [b], Akihiko Noguchi [a]

[a] *Department of Pediatrics, Cardinal Glennon Children's Hospital, Saint Louis University, 1465 S. Grand Blvd., St. Louis, MO 63104, USA*
[b] *Department of Pediatrics, Gifu University, Gifu, Japan*

Received 24 May 2006; accepted 24 May 2006
Available online 11 July 2006

Abstract

Design of efficient treatment strategies for diseases requires clarification of the nature of each mutation causing the disease. In this study, we have investigated three factors to correctly predict the correlation between genotype and phenotype on *N*-acetylgalactosamine-6-sulfate sulfatase (GALNS) gene responsible for one of lysosomal storage diseases, known as mucopolysaccharidosis IVA (MPS IVA); (i) evolutionary conservation of amino acid residues among family proteins, (ii) conservativeness of amino acid changes in *GALNS*, and (iii) structural conservation of amino acid residue. The results showed that (i) the likelihood of a missense variant causing MPS IVA was directly correlated with the level of evolutionary conservation and inversely correlated with conservativeness but not correlated with the structural conservation, (ii) the disease-causative mutations were 9 times more likely to be located on the 'highly conserved' residues than the polymorphisms, (iii) the likelihood of 'non-conservative' amino acid changes in missense mutations was 6.8 times higher than those in the polymorphisms, (iv) the degree of evolutionary conservation was nearly as predictive in phenotype as that of conservativeness of amino acid changes, and (v) the combination of the two factors, evolutionary conservation and conservativeness, provides a better association between missense variants and clinical severity with higher sensitivity (83.5–88.9%) and specificity (71.4–88.3%), than that obtained by either factor alone. These findings suggest that the combination of evolutionary conservation and conservativeness is a useful tool to predict the effect of each mutation on the clinical phenotype and can be applied to the analysis of phenotype/genotype relation in other genetic diseases.

Keywords: Missense variant; Mucopolysaccharidosis IVA; Sulfatase gene family; Conserved amino acid residue; Conservative amino acid change

Introduction

Mucopolysaccharidosis IVA (MPS IVA; Morquio A disease)[2] is an autosomal recessive disorder caused by the deficiency of the lysosomal *N*-acetylgalactosamine-6-sulfate sulfatase (GALNS; E.C.3.1.6.4). GALNS is one of several sulfatases required for the degradation of the glycosaminoglycans, keratan sulfate, and chondroitin-6-sulfate. Phenotypes in MPS IVA vary from a classical form with severe bone dysplasia, short trunk dwarfism, corneal opacity, and a life span of 20–30 years, to an attenuated form with a better quality of life because of mild bone and visceral organ involvement. Investigations of the allelic heterogeneity in MPS IVA were facilitated by the isolation and characterization of the full-length cDNA encoding the human GALNS protein and the genomic GALNS gene [1,2]. To date, about 130 different mutations and seven polymorphisms causing an amino acid change have been

* Corresponding author. Fax: +1 314 577 5398.
E-mail address: tomatsus@slu.edu (S. Tomatsu).
[1] These authors contributed equally to this work.
[2] *Abbreviations used:* MPS IVA, Mucopolysaccharidosis IVA; GALNS, *N*-acetylgalactosamine-6-sulfate sulfatase.

1096-7192/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymgme.2006.05.012

identified. This heterogeneity in GALNS gene mutations accounts for an extensive clinical variability within MPS IVA [3–13].The GALNS enzyme is a member of the sulfatase gene family and 13 different human sulfatase genes are known to date. The GALNS structural model was designed based on homology model of related sulfatases. The tertiary structural model of the human GALNS protein indicated that the severe mutations are associated with the peptide(s) located at the core of the structure leading to destruction of the hydrophobic domain, modification of the packing, or modification of the active site. In contrast, attenuated mutations are mostly associated with peptide abnormalities located on the surface of the GALNS native protein [14].

The frequency and location of missense mutations causing a human genetic disease are non-random. The observed mutational spectrum and the corresponding phenotypic consequences are likely to be determined by the combined effects of (1) the mutation at the DNA level (primary DNA sequence), (2) the level of the evolutionary conservation among the family proteins, (3) the level of conservativeness of the amino acid exchange, and (4) immediate environment within the protein molecule that affects the tertiary protein structure and function by the replaced residue.

The likelihood that a missense variant leads to clinical attention depends largely on two factors; the level of evolutionary conservation and the conservativeness of the resulting amino acid change. The initial studies on a relation between these factors and missense mutations were done in the human factor IX gene, deficiency of which causes hemophilia B [15,16]. Bottema et al. concluded that missense mutations at conserved amino acids led to a disease phenotype more than those at non-conserved residues. Subsequently, Wacey et al. suggested that the level of conservativeness of the amino acid exchange correlated with clinical severity. These two groups have also reported that a high degree of evolutionary conservation of individual amino acid residue predicts functional loss and disease association, and that using few species in evolutionary comparisons overestimates the number of 'conserved residues' [17–20]. The comparative evolutionary methods were also applied to *BRCA1* to identify functionally important amino acid sites by categorizing amino acid sites according to their degree of conservation [21]. However, these studies investigated mutation likelihood by either the level of evolutionary conservation of the exchanged amino acid residue or the level of conservativeness of the resulting amino acid substitution, but not the combined effects. Furthermore, association between the structural conservation of an amino acid residue and the clinical phenotype was also investigated [16].

The GALNS gene was chosen to evaluate the three determinant factors for mutation likelihood and clinical consequence for the following reasons. First, most MPS IVA mutations have been well characterized with regard to genotype/phenotype correlation, and a large proportion (over 70%) of known lesions in the GALNS gene causing MPS IVA derive from missense mutations. Second, seven polymorphisms producing amino acid changes have been identified leading to a better understanding of the difference between disease-causative missense mutations and polymorphic variants. Third, amino acid sequences of GALNS are available for nine species of vertebrates and one of invertebrates. Fourth, the structural model of GALNS has been established [14].

In this study, we examined the association between the clinical phenotype caused by missense variants in *GALNS* and the levels of evolutionary conservation, conservativeness of the amino acid exchange, and structural conservation.

Materials and methods

Subject population

One hundred ten different missense variants (103 missense mutations and seven polymorphic variants) were analyzed primarily for this study. The missense mutations were identified in 190 MPS IVA patients. We categorized the patients according to their overall clinical manifestations and the final height into two groups; severe if the final height is below 125 cm, and attenuated if the final height is above 125 cm (-9SD compared with normal adult height) as described previously [11]. Among the patients studied, 156 were classified as severe, 25 as attenuated, and 9 as undefined. Correlation of each missense variant with the phenotype is based on the following four factors as previously described [13]: (1) homozygosity of the mutation in individual patients, (2) prediction from the most likely structural change in the protein, (3) prediction from *in vitro* expression studies in the GALNS deficient fibroblasts, and (4) presence of a mutant allele, which allows residual enzyme activity in primary fibroblasts or leucocytes, being dominant over an allele that do not permit any activity. As a result, 63 missense variants were defined as severe while 30 were as attenuated. Ten mutations were not defined with the specific phenotype by the above criteria. All seven polymorphisms were observed in a homozygote in normal population. In addition, eight nonsense mutations were also described for comparison. A frequency of mutations and polymorphisms was described previously [12,13].

Levels of evolutionary conservation of the amino acid residue in GALNS

The GALNS amino acid sequences of human, dog, cow, pig, mouse, rat, zebrafish, pufferfish, frog, chicken, and sea urchin were retrieved from GenBank. The multiple alignment of the amino acid sequences of GALNSs was performed by using the CLUSTAL-W alignment program [22], and manually adjusted using GeneDoc program [23]. Each missense, nonsense, or polymorphic amino acid change is shown with a defined clinical phenotype. Fig. 2 shows the partial alignment between 62 and 146 amino acid residues of human GALNS. To investigate the relationship between the level of evolutionary conservation of amino acid residues and clinical severity (severe, attenuated, or polymorphic), each amino acid residue in the GALNS protein molecule was classified into four classes ('conserved', 'vertebrate-specific', 'mammal-specific', and 'non-conserved') by the extent of homology between the 11 species of GALNSs. Furthermore, the 'conserved' plus 'vertebrate-specific' amino acid residues were grouped as 'highly conserved', while the 'mammal-specific' and 'non-conserved' were defined as 'less-conserved'.

An amino acid residue in GALNS protein was 'conserved' if it was identical or similar among the eleven species of GALNSs. The residue was 'vertebrate-specific' if it was identical or similar among the 10 vertebrate species of GALNSs. If a residue was identical or similar among the six mammalian species of GALNSs, it was classified as 'mammal-specific'. If a residue was neither identical nor similar among the species of GALNSs, it was classified as 'non-conserved'. The criteria for