GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

In silico analysis of mutations occurring in the protein N-acetylgalactosamine-6-sulfatase (GALNS) and causing mucopolysaccharidosis IVA

E.R. Tamarozzi, E. Torrieri, E.P. Semighini and S. Giuliatti

Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, SP, Brasil

Corresponding author: E.R. Tamarozzi
E-mail: Elvira.tamarozzi@gmail.com

Genet. Mol. Res. 13 (4): 10025-10034 (2014)
Received April 29, 2014
Accepted August 7, 2014
Published November 28, 2014
DOI http://dx.doi.org/10.4238/2014.November.28.7

ABSTRACT. The goals were to analyze and characterize the secondary structure, regions of intrinsic disorder and physicochemical characteristics of three classes of mutations described in the enzyme N-acetylgalactosamine-6-sulfatase that cause mucopolysaccharidosis IVA: missense mutations, insertions and deletions. All mutations were compared to wild-type enzyme, and the results showed that with 25 of 129 missense mutations secondary structure was maintained and that 104 mutations showed minor changes, such as an increase or decrease in the size of the elements. The secondary structure of all insertions and deletions introduced important changes, such as increase in the number and size of elements. The results obtained from intrinsic disorder analysis revealed that missense mutations caused no alterations. However, the insertions and deletions led to major regions of intrinsic disorder. The physicochemical characteristics of the amino acids found in missense mutations revealed unchanged characteristics in 32 of the 129 mutations. However, the remainder had changes that could lead to a modification of tertiary structure.

The results proved that it was feasible and necessary to obtain the three-dimensional structure of the enzyme with its mutants to better understand the change in function.

Key words: Morquio syndrome; Bioinformatics; Protein structure

INTRODUCTION

Intrinsically disordered proteins perform several important roles in cellular activities, and they are related to a large number of diseases (Mizianty et al., 2011). The traditional structure-function relationship approach has, as one of its principles, the concept that the biological function of a protein is critically dependent on a well-defined conformational structure (Dunker et al., 2002). The deficiencies in lysosomal enzymes responsible for the catabolism of biomolecules lead to the accumulation of undegraded substrates in a wide range of human tissues and, eventually, to the so-called deposit diseases (Neufeld and Muenzer, 2001). N-acetylgalactosamine-6-sulfatase (GALNS) is the enzyme responsible for the first step in the degradation process of the glycosaminoglycans (GAG) called keratan sulfate (KS) and chondroitin-6-sulfate (C6S). It hydrolyzes the bond between GAG and the sulfate group linked to the 6th carbon of the saccharides, and its deficiency leads to the accumulation of KS and C6S in lysosomes, causing a cellular dysfunction known as mucopolysaccharidosis (MPV) IV, or Morquio syndrome (MS) (Pajares et al., 2012; Rivera-Colón et al., 2012; Kubaski et al., 2013).

MS is an autosomal recessive disease, acquired from healthy heterozygous parents carrying a mutation in band q24.3 of chromosome 16, a mutation that was discovered by *in situ* hybridization assays (Tomatsu et al., 2005). SM is characterized by severe osseous deformities and extra skeletal changes, such as corneal opacity, hepatomegaly, cardiac valve lesions, prognatism, diastema, thinning tooth enamel, and auditory changes that vary from conductive to sensorineural dysacusis. (Montaño et al., 2007; Couprie et al., 2010; Leal et al., 2010; Tomatsu et al., 2011; Hendriksz et al., 2013).

MS has two phenotypes related to its severity: a mild form, which generally allows the patient to have a normal life, and a severe form, which usually leads to the patient's death before the second decade of life.

Genetic analysis in different populations has shown a heterogeneity in the q24.3 mutation. However, the correlation between genotypes and phenotypes is not yet clearly understood (Tomatsu et al., 2011; Rivera-Colón et al., 2012; Pajares et al., 2012; Kubaski et al., 2013).

It is difficult to estimate the prevalence of MS, due to shortages of basic population studies and epidemiological data, which is estimated to be between 1 in 40,000 and 1 in 50,000 births (Northover et al., 1996; Nelson, 1997; Rivera-Colón et al., 2012). In Brazil, only the state of Rio Grande do Sul has data on the incidence of MS, showing 1 in every 275,000 live births (Schwartz et al., 2008).

Up to the submission of this paper, there was no known cure for MS. Current treatment consists of a multidisciplinary management with psychological, physiotherapeutic, nutritional, medical, and surgical interventions. These interventions are palliative and preventive and aim to delay the progress of the disease (Algahim and Almassi, 2013).

The most often reported mutations of GALNS enzyme that cause MS are classified in three categories: changes in the catalytic site, changes in the hydrophobic core and changes in

enzyme surface (Tomatsu et al., 2004; Rivera-Colón et al., 2012).

Tomatsu et al. (2005) reported a total of 148 mutations, with 103 of them being missense mutations, a number recently revised by Rivera-Colón et al. (2012) to 129 missense mutations. Since the tertiary structure of the GALNS enzyme is known (Rivera-Colón et al., 2012), the search for possible regions of intrinsic disorder in the mutant protein has substantial importance to enable protein modeling and virtual screening assays of these mutants, allowing the discovery of molecules with theoretic binding potential to the enzyme, and eventual pharmacological reactivation of the affected enzyme.

The purpose of this study was to perform *in silico* analysis of possible regions of intrinsic disorder (ID) and to compare the secondary structure of the wild-type protein with the 161 known (129 missense, 5 insertions, and 27 deletions) mutations, to determine if the increase in ID is responsible for the protein's loss of function.

MATERIAL AND METHODS

The genetic mutation data of the GALNS gene were described by Terzioglu (2002), Tomatsu et al. (2004), Tomatsu et al. (2005) and Rivera-Colón et al. (2012).

The FASTA sequence of GALNS has 522 amino acids and was downloaded in FASTA format from the NCBI public database, available at http://www.ncbi.nlm.nih.gov/protein/P34059.

The comparative analysis of chemical characteristics of wild type GALNS and the missense mutants followed the technique described by Nelson and Cox (2009).

The ID regions of the wild type and mutant proteins were predicted by the softwares meta Protein DisOrder prediction System - metaPrDOS (Ishida and Kinoshita, 2008), Prediction of Intrinsically Unstructured Proteins - IUPred (Dosztányi et al., 2005), Disorder Content predictor - DisCon (Mizianty et al., 2011), POODLE-S (Shimizu et al., 2005) and Prediction of Protein Binding Regions in Disordered Proteins - ANCHOR (Dosztányi et al., 2009).

The PSIPRED software (Buchan et al., 2013) was used to predict the secondary structures of the mutants.

RESULTS

All 161 mutations (Terzioglu et al., 2002; Tomatsu et al., 2004, 2005; Rivera-Colón et al., 2012) were individually analyzed, and the results were compared to the data for wild-type GALNS.

The results of predicted secondary structures, described from PSIPRED, revealed that wild-type GALNS has 9 α-helices and 9 β-sheets. It also revealed small structural changes at α-helices and/or β-sheets in 104 of the 129 missense mutations when compared to wild-type GALNS (Table 1). On the other hand, the 27 deletions and 5 insertions showed substantial changes in the size and number of α-helices and β-sheets in comparison with wild-type GALNS (Table 2).

The next step was the prediction of the ID regions, where there was no pronounced discrepancy between the results of the five software used to predict the ID regions, considering 5% as the margin of error.

The results showed that wild-type GALNS has an average of less than 2.3% (about 10 aminoacid residues) of ID in its whole length. Besides the differences between the methods

Table 1. GALNS mutations and its effects on secondary structures and intrinsic disorder (ID).

Mutation type	Protein region	Amino acid changed and mutation place	Average ID of the mutation	Secondary structure change	Structure changed (sheet/helix)
Missense	Surface	E 185 G	<2.4%	Yes	Both
Missense	Surface	N 204 A	<2.4%	Yes	Both
Missense	Surface	L 395 V	<2.4%	Yes	Both
Missense	Surface	H 398 D	<2.4%	Yes	Both
Missense	Surface	G 421 E	<2.4%	Yes	Both
Missense	Surface	Q 422 K	<2.4%	Yes	Both
Missense	Surface	F 452 I	<2.4%	Yes	Both
Missense	Surface	F 452 L	<2.4%	Yes	Both
Missense	Surface	S 53 F	<2.4%	Yes	Sheet
Missense	Surface	D 60 N	<2.4%	Yes	Sheet
Missense	Surface	G 155 E	<2.4%	Yes	Sheet
Missense	Surface	G 155 R	<2.4%	Yes	Sheet
Missense	Surface	F 167 V	<2.4%	Yes	Sheet
Missense	Surface	D 171 A	<2.4%	Yes	Sheet
Missense	Surface	R 259 Q	<2.4%	Yes	Sheet
Missense	Surface	S 295 F	<2.4%	Yes	Sheet
Missense	Surface	L 369 P	<2.4%	Yes	Sheet
Missense	Surface	W 409 S	<2.4%	Yes	Sheet
Missense	Surface	P 484 S	<2.4%	Yes	Sheet
Missense	Surface	V 488 M	<2.4%	Yes	Sheet
Missense	Surface	R 253 W	<2.4%	Yes	Helix
Missense	Surface	N 487 S	<2.4%	Yes	Helix
Missense	Surface	R 61 W	<2.4%	No	
Missense	Surface	G 247 D	<2.4%	No	
Missense	Surface	A 351 V	<2.4%	No	
Missense	Surface	P 357 L	<2.4%	No	
Missense	Surface	R 361 G	<2.4%	No	
Missense	Surface	L 366 F	<2.4%	No	
Missense	Surface	R 376 Q	<2.4%	No	
Missense	Surface	D 388 N	<2,.4%	No	
Missense	Surface	L 395 P	<2.4%	No	
Missense	Surface	S 470 P	<2.4%	No	
Missense	Active Site	C 79 Y	<2.4%	Yes	Both
Missense	Active Site	H 142 R	<2.4%	Yes	Both
Missense	Active Site	S 80 L	<2.4%	Yes	Sheet
Missense	Active Site	H 236 D	<2.4%	Yes	Sheet
Missense	Active Site	N 289 S	<2.4%	No	
Missense	Active Site	K 310 N	<2.4%	No	
Missense	Hydrophobic Core	G 96 C	<2.4%	Yes	Both
Missense	Hydrophobic Core	F 97 V	<2.4%	Yes	Both
Missense	Hydrophobic Core	A 107 T	<2.4%	Yes	Both
Missense	Hydrophobic Core	S 135 R	<2.4%	Yes	Both
Missense	Hydrophobic Core	V 138 A	<2.4%	Yes	Both
Missense	Hydrophobic Core	W 141 C	<2.4%	Yes	Both
Missense	Hydrophobic Core	W 141 R	<2.4%	Yes	Both
Missense	Hydrophobic Core	P 151 S	<2.4%	Yes	Both
Missense	Hydrophobic Core	P 163 H	<2.4%	Yes	Both
Missense	Hydrophobic Core	H 166 Q	<2.4%	Yes	Both
Missense	Hydrophobic Core	A 291 T	<2.4%	Yes	Both
Missense	Hydrophobic Core	L 36 P	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 41 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 42 E	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 47 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 66 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 69 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 77 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	T 88 I	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 94 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 94 G	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 96 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	Q 111 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 116 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 125 L	<2.4%	Yes	Sheet

Continued on next page

Table 1. Continued.

Mutation type	Protein region	Amino acid changed and mutation place	Average ID of the mutation	Secondary structure change	Structure changed (sheet/helix)
Missense	Hydrophobic Core	H 150 Y	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 151 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 159 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 162 F	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	N 164 T	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 168 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 179 H	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 179 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 203 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 230 G	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	V 239 F	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	E 260 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 284 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 287 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 290 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 291 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 301 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	L 307 P	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 309 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	T 312 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 316 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 318 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 324 E	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 325 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 340 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 341 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 343 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 343 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	D 344 N	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 346 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 380 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 380 T	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 386 H	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 391 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 392 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	H 401 Y	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	E 450 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	D 233 N	<2.4%	Yes	Helix
Missense	Hydrophobic Core	R 90 W	<2.4%	No	
Missense	Hydrophobic Core	I 113 F	<2.4%	No	
Missense	Hydrophobic Core	G 139 S	<2.4%	No	
Missense	Hydrophobic Core	F 156 S	<2.4%	No	
Missense	Hydrophobic Core	P 179 S	<2.4%	No	
Missense	Hydrophobic Core	A 257 T	<2.4%	No	
Missense	Hydrophobic Core	T 312 A	<2.4%	No	
Missense	Hydrophobic Core	D 344 E	<2.4%	No	
Missense	Hydrophobic Core	L 345 P	<2.4%	No	
Missense	Hydrophobic Core	L 352 P	<2.4%	No	
Missense	Hydrophobic Core	R 386 C	<2.4%	No	
Missense	Hydrophobic Core	N 407 H	<2.4%	No	
Missense	Hydrophobic Core	M 494 V	<2.4%	No	
Missense	Undescribed	M 1 V	<2.4%	Yes	Both
Missense	Undescribed	W 10 X	<2.4%	Yes	Both
Missense	Undescribed	L 15 M	<2.4%	Yes	Both
Missense	Undescribed	Q 111 X	<2.4%	Yes	Both
Missense	Undescribed	K 129 X	<2.4%	Yes	Both
Missense	Undescribed	Q 148 X	<2.4%	Yes	Both
Missense	Undescribed	Q 211 X	<2.4%	Yes	Both
Missense	Undescribed	W 230 X	<2.4%	Yes	Both
Missense	Undescribed	W 325 X	<2.4%	Yes	Both
Missense	Undescribed	Q 338 X	<2.4%	Yes	Both
Missense	Undescribed	Q 374 X	<2.4%	Yes	Both
Missense	Undescribed	Q 422 X	<2.4%	Yes	Both
Missense	Undescribed	Q 473 X	<2.4%	Yes	Both
Missense	Undescribed	G 23 R	<2.4%	Yes	Sheet

Table 2. Secondary structure changes and intrisic disorder (ID) of the 27 deletions and 5 insertions known in GALNS.

Mutation type	Mutation place	No. of deleted or inserted nucleotides	Rise greater than 2.5% of ID	ID percentual of the mutation	Major changes at secondary structures
Deletion	334	1	Yes	32.0%	Yes
Deletion	489	1	Yes	27.0%	Yes
Deletion	498	1	Yes	27.0%	Yes
Deletion	605	1	Yes	32.0%	Yes
Deletion	708	1	Yes	25.0%	Yes
Deletion	874	1	Yes	20.0%	Yes
Deletion	1168		Yes	15.0%	Yes
Deletion	1319	1	Yes	7.5%	Yes
Deletion	121 to 244	124	Yes	42.0%	Yes
Deletion	1213 to 1219	6	No	2.2%	Yes
Deletion	1290 to 1291	2	Yes	14.5%	Yes
Deletion	154 to 165	1	Yes	38.0%	Yes
Deletion	209 to 220	12	No	2.2%	Yes
Deletion	235 to 236	2	Yes	41.0%	Yes
Deletion	49 to 55	7	Yes	42.0%	Yes
Deletion	532 to 533	2	Yes	42.0%	Yes
Deletion	833 to 859	26	Yes	26.0%	Yes
Deletion	853 to 855	3	No	2.5%	Yes
Deletion	976 to 978	3	No	2.5%	Yes
Deletion	106 to 111	6	No	2.5%	Yes
Deletion	655	1	Yes	38.0%	Yes
Deletion	600	1	Yes	38.0%	Yes
Deletion	389	1	Yes	42.0%	Yes
Deletion	763	1	Yes	35.0%	Yes
Deletion	929	1	Yes	35.0%	Yes
Deletion	850 to 852	3	No	2.5%	Yes
Deletion	1343 to 1344	2	Yes	16.0%	Yes
Insertion	556	1	Yes	41%	Yes
Insertion	1232-1233	1	Yes	18%	Yes
Insertion	1177-1178	1	Yes	15.0%	Yes
Insertion	1355-1356	1	Yes	7.0%	Yes
Insertion	501-502	1	Yes	42.0%	Yes

of ID prediction, metaPrDOS, IUPred, DisCon, POODLE-S and ANCHOR, all the results of wild-type GALNS were consistent with each other.

The 129 missense mutations analyzed were grouped into four groups: 6 mutations at the catalytic site residues, 32 at the protein surface residues, 76 mutations at the protein hydrophobic core and 14 mutations at undescribed regions. All results were similar, without any considerable anomaly between the methods, with up to 2.4% of an average predicted ID for these mutations (Table 1), and also similar to those from the secondary structure predictions.

However, at the 27 deletions and 5 insertions, predicted ID varied from 2.2 to 42% (Table 2).

DISCUSSION

Tomatsu et al. (2005) and Rivera-Colón et al. (2012) showed that missense mutations [also called non synonymous substitution (Zhang et al., 2010, 2012)] are the most common cause of MPS IVA. It causes amino acid changes due to single nucleotide polymorphisms, which can drastically change protein structure and function. The specific mechanism that causes these mutations remains unknown. Yue and Moult (2006) and Care et al. (2007) reported that non synonymous substitution mutations have an impact on the structural stability of the protein, altering its function. These data led us to analyze the secondary structures of the wild-type enzyme and its mutants, searching for the presence of regions with ID in the proteins.

According to Gnad et al. (2013), depending on its position and nature of the amino acid substitution of these mutations, it can cause a variety of changes of variable proportions in protein structure, affecting GALNS physiological function.

The missense mutations present in GALNS can cause changes in its active site, hydrophobic core and surface (Rivera-Colón et al., 2012). Zhang et al. (2011, 2012) reported that any change that affects the geometry of the active site (mutation that can be outside of the catalytic site) of a protein can, thus, affect its function. As our results showed, the changes in the secondary structures of missense mutant GALNS can explain the lower or inexistent enzymatic activity of these proteins. The analysis of the chemical characteristics of substituted amino acids has also showed that, in most of the missense mutations, they were modified, explaining then the predicted changes in the secondary structures.

Valdés et al. (2013) outlined that the ID, under physiological conditions, is common in a wide range of proteins. This phenomenon plays a key role in the biological activities of proteins that have a diverse range of molecular recognition patterns. To understand the role of ID in protein function, it is important to determine if the ID is part of its physiological role or, if the disorder is caused by mutations, leading to diseases such as synucleinopathies (Dill et al., 2007; Uversky et al., 2009).

It was expected to see ID caused by missense mutations in MPS IV-related GALNS. However, the mutant protein responsible for MPS IVA did not show ID regions, just like the wild-type protein, corroborating the findings of Rivera-Colón et al. (2012), showing GALNS as a very specific and conserved protein in vertebrates.

Protein stability is a key feature to protein function (Ye et al., 2006; Zhang et al., 2012), and a missense mutation may affect, besides the secondary structure, its stability, leading to misfolding (Dobson, 2003; Koukouritaki et al., 2007; Zhang et al., 2012). Our findings showed that, unexpectedly, even in missense mutations, without ID regions, when there are no changes in 19.4% of the secondary structures and in 25.6% of the chemical characteristics of

the amino acids, the function of GALNS is still suppressed in patients with MPS IVA. In the remaining part that showed minor changes in secondary structure (80.4%) and changes in the chemical characteristics of the amino acids (74.4%), no ID regions were observed. However, these changes may explain the suppression of protein function, since these variations can cause misfolding and also change the pattern of substrate recognition, which, in this case, is very specific to KS and C6S.

The missense mutations are not the only cause of MPS IVA. Maki (2002) pointed out that, in general terms, mutations involving deletions and insertions are less frequent than missense mutations. In analyzing the results of ID regions caused by insertions and deletions and the prediction of their secondary structures, we observed an increase in ID regions and wide changes in secondary structures, in contrast to that observed in missense mutations, but very expected, and that is probably the reason for the loss of enzyme function, since the whole amino acid sequence changes after a nucleotide deletion or insertion.

The prediction of secondary structures and the search for ID regions through *in silico* analysis of wild-type GALNS and its mutations led us to two different results. The results of the missense mutations analysis led us to infer, by observing their secondary structures conserved in more than 99% of its length, that it is possible to model their tertiary structures from wild-type GALNS. These results are reinforced by Uversky et al. (2009), who showed that proteins without ID have a more stable geometry at the active site in comparison to disordered proteins, increasing the chances of finding ligands to GALNS. On the other hand, the analysis of deletions and insertions showed it to be different from that obtained for missense mutations, due to major changes in the primary and secondary structure. These results do not discard the modeling of mutant GALNS and are very important in the understanding of MPS IVA and its phenotypic variation process.

CONCLUSION

Of all the 161 mutations analyzed, 80% were missense and did not show a significant percentage of ID increase when compared to wild-type GALNS. The 27 deletions and 5 insertions showed a high increase in ID in most of its structures. Secondary structure analysis revealed that 16% of the mutations, all missense, had no changes, 64% of the mutations, also all missense, showed small changes and 20% of the mutations (insertions and deletions) showed wide changes in secondary structures in comparison to wild-type GALNS. The analysis of the chemical features of the mutated amino acids showed that 25% of them were replaced by homologous amino acids.

On the basis of these data, it can be concluded that it is possible and viable to use wild-type GALNS to model the structures of 80.1% of the mutations and then to use these models to understand how these mutations affect protein catalytic activity.

REFERENCES

Algahim MF and Almassi GH (2013). Current and emerging management options for patients with Morquio A syndrome. *Ther. Clin. Risk Manag.* 9: 45-53.

Buchan DW, Minneci F, Nugent TC, Bryson K, et al. (2013). Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* 41: W349-W357.

Care MA, Needham CJ, Bulpitt AJ and Westhead DR (2007). Deleterious SNP prediction: be mindful of your training data! *Bioinformatics* 23: 664-672.

Couprie J, Denis P, Guffon N, Reynes N, et al. (2010). Manifestations ophtalmologiques de la maladie de Morquio. *J. Français d'Ophtalmol.* 33: 617-622.

Dill KA, Ozkan SB, Weikl TR, Chodera JD, et al. (2007). The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.* 17: 342-346.

Dobson CM (2003). Protein folding and misfolding. *Nature* 426: 884-890.

Dosztányi Z, Csizmók V, Tompa P and Simon I (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21: 3433-3434.

Dosztányi Z, Meszaros B and Simon I (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25: 2745-2746.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, et al. (2002). Intrinsic disorder and protein function. *Biochemistry* 41: 6573-6582.

Gnad F, Baucom A, Mukhyala K, Manning G, et al. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* 14 (Suppl 3): S7.

Hendriksz CJ, Al-Jawad M, Berger KI, Hawley SM, et al. (2013). Clinical overview and treatment options for non-skeletal manifestations of mucopolysaccharidosis type IVA. *J. Inherit. Metab. Dis.* 36: 309-322.

Ishida T and Kinoshita K (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 24: 1344-1348.

Koukouritaki SB, Poch MT, Henderson MC, Siddens LK, et al. (2007). Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants. *J. Pharmacol. Exp. Ther.* 320: 266-273.

Kubaski F, Brusius-Facchin AC, Palhares HM, Balarin MA, et al. (2013). Identification of a novel missense mutation in Brazilian patient with a severe form of mucopolysaccharidosis type IVA. *Gene* 517: 112-115.

Leal GN, de Paula AC, Leone C and Kim CA (2010). Echocardiographic study of paediatric patients with mucopolysaccharidosis. *Cardiol. Young* 20: 254-261.

Maki H (2002). Origins of spontaneous mutations: specificity and directionality of base-substitution, frame shift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* 36: 279-303.

Mizianty MJ, Zhang T, Xue B, Zhou Y, et al. (2011). *In-silico* prediction of disorder content using hybrid sequence representation. *BMC Bioinformatics* 12: 245.

Montaño AM, Tomatsu S, Gottesman GS, Smith M, et al. (2007). International Morquio A Registry: clinical manifestation and natural course of Morquio A disease. *J. Inherit. Metab. Dis.* 30: 165-174.

Nelson DL and Cox MM (2009). Lehninger Principles of Biochemistry. 5th edn. W.H. Freeman Publishers, New York.

Nelson J (1997). Incidence of the mucopolysaccharidoses in Northern Ireland. *Hum. Genet.* 101: 355-358.

Neufeld EF and Muenzer J (2001). The Mucopolysaccharidoses. In: The Metabolic & Molecular Bases of Inherited Disease. 8th edn. Vol. 3. (Scriver CR, Beaudet AL, Sly WS and Valle D, eds.). McGraw-Hill, New York.

Northover H, Cowie RA and Wraith JE (1996). Mucopolysaccharidosis type IVA (Morquio syndrome): a clinical review. *J Inherit. Metab. Dis.* 19: 357-365.

Pajares S, Alcalde C, Couce ML, Del TM, et al. (2012). Molecular analysis of mucopolysaccharidosis IVA (Morquio A) in Spain. *Mol. Genet. Metab.* 106: 196-201.

Rivera-Colón Y, Schutsky EK, Kita AZ and Garman SC (2012). The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. *J. Mol. Biol.* 423: 736-751.

Schwartz I, Martins TF and Federhen A (2008). A Rede MPS Faz Quatro Anos: Um Panorama Geral sobre a Epidemiologia das MPS no Brasil. In: Anais do XX Congresso Brasileiro de Genética Médica, Brasil.

Shimizu K, Muraoka Y, Hirose S and Noguchi T (2005). Feature Selection Based on Physicochemical Properties of Redefined N-term Region and C-term Regions for Predicting Disorder, Proc of IEEE CIBCB, 262-267.

Terzioglu M, Tokatli A, Coskun T and Emre S (2002). Molecular analysis of Turkish mucopolysaccharidosis IVA (Morquio A) patients: identification of novel mutations in the N-acetylgalactosamine-6-sulfate sulfatase (GALNS) gene. *Hum. Mutat.* 20: 477-478.

Tomatsu S, Nishioka T, Montaño AM, Gutierrez MA, et al. (2004). Mucopolysaccharidosis IVA: identification of mutations and methylation study in GALNS gene. *J. Med. Genet.* 41: e98.

Tomatsu S, Montaño AM, Nishioka T, Gutierrez MA, et al. (2005). Mutation and polymorphism spectrum of the GALNS gene in mucopolysaccharidosis IVA (Morquio A). *Hum. Mutat.* 26: 500-512.

Tomatsu S, Montaño AM, Oikawa H, Smith M, et al. (2011). Mucopolysaccharidosis type IVA (Morquio A disease): clinical review and current treatment. *Curr. Pharm. Biotechnol.* 12: 931-945.

Uversky VN, Oldfield CJ, Midic U, Xie H, et al. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* 10 (Suppl 1): S7.

Valdés JJ, Schwarz A, Cabeza de Vaca I and Calvo E (2013). Tryptogalinin is a tick kunitz serine protease inhibitor with a unique intrinsic disorder. *PLoS One* 8: 62562.

Ye Y, Li Z and Godzik A (2006). Modeling and Analyzing Three-Dimensional Structures of Human Disease Proteins. Proceedings of the Pacific Symposium on Biocomputing, Hawaii, 439-450.

Yue P and Moult J (2006). Identification and analysis of deleterious human SNPs. *J. Mol. Biol.* 356: 1263-1274.

Zhang Z, Teng S, Wang L, Schwartz CE, et al. (2010). Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum. Mutat.* 31: 1043-1049.

Zhang Z, Norris J, Schwartz C and Alexov E (2011). *In silico* and *in vitro* investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS One* 6: e20373.

Zhang Z, Miteva MA, Wang L and Alexov E (2012). Analyzing effects of naturally occurring missense mutations. *Comput. Math. Methods Med.* 2012: 1-15.