Published in final edited form as:

Hum Mutat. 2013 November ; 34(11): 1558–1567. doi:10.1002/humu.22429.

Position of glycine substitutions in the triple helix of *COL6A1*, *COL6A2*, and *COL6A3* is correlated with severity and mode of inheritance in collagen VI myopathies

Russell J. Butterfield¹, A. Reghan Foley², Jahannaz Dastgir³, Stephanie Asman¹, Diane M. Dunn⁵, Yaqun Zou³, Ying Hu³, Kevin M. Flanigan⁴, Kathryn J. Swoboda¹, Thomas L. Winder⁶, Robert B. Weiss^{5,1}, and Carsten G. Bönnemann³

¹University of Utah, Departments of Pediatrics and Neurology, Salt Lake City, UT, USA

²Dubowitz Neuromuscular Centre, UCL Institute of Child Health and Great Ormond Street Hospital for Children, London, UK

³Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA

⁴Center for Gene Therapy, Nationwide Children's Hospital, Columbus, OH, USA

⁵University of Utah, Department of Human Genetics, Salt Lake City, UT, USA

⁶Prevention Genetics, Marshfield, WI, USA

Abstract

Glycine substitutions in the conserved Gly-X-Y motif in the triple helical domain of collagen VI are the most commonly identified mutations in the collagen VI myopathies including Ullrich congenital muscular dystrophy, Bethlem myopathy, and intermediate phenotypes. We describe clinical and genetic characteristics of 97 individuals with glycine substitutions in the triple helical domain of *COL6A1*, *COL6A2*, or *COL6A3* and add a review of 97 published cases, for a total of 194 cases. Clinical findings include severe, intermediate, and mild phenotypes even from patients with identical mutations. Intermediate phenotypes were most common, accounting for almost half of patients, emphasizing the importance of intermediate phenotypes to the overall phenotypic spectrum. Glycine substitutions in the triple helical domain are heavily clustered in a short segment N-terminal to the 17th Gly-X-Y triplet, where they are acting as dominants. The most severe cases are clustered in an even smaller region including Gly-X-Y triplets 10 to 15, accounting for only 5% of the triple helical domain. Our findings suggest that clustering of glycine substitutions in the N-terminal region of collagen VI is not based on features of the primary sequence. We hypothesize that this region may represent a functional domain within the triple helix.

Correspondence: Carsten Bönnemann, MD, Bldg 35 Porter NRC, Room 2A-116, 35 Convent Drive MSC 3705, Bethesda, Maryland 20892-3705, Phone: 301-594-5496, Fax: 301-480-3365, carsten.bonnemann@nih.gov.

Supporting Information for this preprint is available from the *Human Mutation* editorial office upon request (humu@wiley.com)

Disclosure statement: The authors declare no competing financial interests

Keywords

collagen VI; Ullrich congenital muscular dystrophy; Bethlem myopathy; genotype-phenotype correlation

INTRODUCTION

The collagen VI myopathies, Ullrich congenital muscular dystrophy (UCMD) and Bethlem myopathy (BM) are among the most common congenital muscular dystrophies and are characterized by distal joint laxity and a combination of distal and proximal joint contractures (Clement, et al., 2012; Okada, et al., 2007). In UCMD (MIM# 254090), progressive weakness is manifest in the neonatal period or early childhood, frequently resulting in early loss of ambulation (Ullrich, 1930). In BM (MIM# 158810) weakness begins mid-childhood or early adolescence, but progression is slow and ambulation is retained into adulthood (Jobsis, et al., 1996). UCMD and BM had been considered distinct clinical and genetic entities until both were linked to mutations in the genes encoding collagen VI (*COL6A1*, MIM# 120220; *COL6A2*, MIM# 120240; *COL6A3*, MIM# 120250) (Camacho Vanegas, et al., 2001; Jobsis, et al., 1996). UCMD and BM phenotypes are thought to represent the ends of a clinical spectrum that includes intermediate phenotypes of variable severity. A myosclerosis myopathy phenotype (Merlini, et al., 2008) (MIM# 255600) and a limb-girdle muscular dystrophy phenotype (Scacheri, et al., 2002) have also been described.

Collagen VI is a ubiquitous non-fibrillar collagen composed of three chains, $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ organized into a network of microfibrils important in anchoring the basement membrane to the extracellular matrix (ECM) (Kuo, et al., 1997). Each chain contains a comparatively short triple helical (TH) domain with repeating Gly-X-Y subunits flanked by large globular von Willebrand factor type A domains (Fig. 1) (Chu, et al., 1989; Chu, et al., 1990a). Cultured skin fibroblasts from UCMD and BM patients have decreased matrix deposition and poor localization of collagen VI to the basement membrane (Hicks, et al., 2008).

Genotype/phenotype associations in the collagen VI myopathies have been difficult to assess due to the considerable clinical and genetic heterogeneity, including presentations of UCMD, BM, and intermediate phenotypes in patients with similar mutations (Baker, et al., 2007; Baker, et al., 2005; Brinas, et al., 2010; Lampe, et al., 2005; Reed, et al., 2005). Missense mutations involving the conserved glycine residue in the repeated Gly-X-Y motif in the TH domain are a common pathogenic mutation in both collagen VI myopathies and in disorders of other collagens. In the collagen VI genes, such glycine substitutions account for 30% of known pathogenic alleles (Brinas, et al., 2010; Lampe and Bushby, 2005). In contrast to glycine substitutions in the TH domain of other collagens (Dang and Murrell, 2008; Marini, et al., 2007; Pescucci, et al., 2003), glycine substitutions in collagen VI appear to cluster in the N-terminal end of the TH domain (Pace, et al., 2008). Substitution of a glycine residue or exon skipping in the N-terminal end of the TH domain does not appear to disrupt the formation of triple helical monomers, but it does disrupt higher order assembly,

thus resulting in a dominant negative mode of action (Lamande, et al., 2002; Pace, et al., 2008). It has been suggested that glycine substitutions in a critical region including Gly-X-Y triplets 10-15 in the N-terminal triple helix are correlated with a more severe disruption of assembly and with increased clinical severity (Pace, et al., 2008).

Here, we describe clinical and genetic characteristics of 97 individuals with glycine substitutions in the TH domain of the collagen VI $\alpha 1$ (VI), $\alpha 2$ (VI), and $\alpha 3$ (VI) chains, and add a review of 97 published cases, for a total of 194 cases. Clinical findings in these patients include severe UCMD, intermediate, and mild BM phenotypes. Identical glycine substitutions are associated with both severe and mild phenotypes. In all 3 chains, glycine substitutions are heavily clustered in a short segment of the triple helical domain N-terminal to the 17th Gly-X-Y triplet (TH17). The most severe cases are clustered in the critical tetramer assembly region including Gly-X-Y triplets 10 to 15 accounting for only 5% of the length of the triple helical domain.

SUBJECTS AND METHODS

Patients

Ninety-seven individuals from 83 families with missense mutations resulting in glycine substitutions in the TH domain of the *COL6A1*, *COL6A3*, and *COL6A3* genes were identified from patients undergoing clinical genetic testing performed at the University of Utah (Salt Lake City, UT) and Prevention Genetics (Marshfield, WI), or who were seen clinically at the University of Utah (RB), Children's Hospital of Philadelphia (ARF and CGB), or National Institutes of Health, Neuromuscular and Neurogenetic Disorders of Childhood Section (JD and CGB). Patients were enrolled in our collaborative collagen VI myopathy project and written informed consent was obtained according to the ethics committees of the participating institutions. Phenotype data were collected from medical records and patient questionnaires with an emphasis on major motor events such as initiation and loss of ambulation (Nadeau, et al., 2009). We collected data on pulmonary function when it was available, but due to the heterogeneity of the patient population, we did not have sufficient data for correlation with respiratory outcomes. We also conducted an extensive literature search and identified 97 additional patients from 65 families with published glycine substitutions in the TH domain. Phenotype of individuals in published reports was collected with a similar focus on motor function and in particular ambulation.

Phenotype

Patients were divided into groups based on clinical severity and progression. The early severe (ES) UCMD category includes patients who never achieved independent ambulation (Brinas, et al., 2010). The typical UCMD category includes patients who achieved independent ambulation but subsequently lose ambulation by 12 (an average age of 10), or who remain ambulatory indoors only. This 'typical' UCMD group is comparable with the moderate progressive UCMD patients reported by Brinas *et al.* (Brinas, et al., 2010). Patients under 12 years of age at the time of evaluation were classified as UCMD if their first wheelchair use was before age 8 or they started nighttime noninvasive ventilation before age 10. The intermediate (INT) UCMD category includes patients who have

remained ambulant beyond 12 years of age but have significant motor impairment with loss of independent ambulation by adulthood. Patients under 12 years of age were classified as INT if they had modest gait impairment and infrequent use of assistive devices in childhood. Bethlem myopathy patients demonstrated independent ambulation into adulthood without significant gait impairment. Patients less than 18 years of age were classified as BM if they had joint contractures and minimal motor impairment or if a BM phenotype was predicted by family history. We were able to determine a phenotype classification in 51 of 97 cases in our cohort and in 52 of 97 published cases based on clinical data presented in the respective publications (Supp. Tables S1, S2).

Mutation analysis

Sequence analysis of the *COL6A1*, *COL6A2*, and *COL6A3* genes was performed from genomic DNA as a clinical genetic test in CLIA-certified laboratories (University of Utah, Salt Lake City, UT and Prevention Genetics, Marshfield, WI) by Sanger sequencing methods, including SCAIP sequencing (Lampe, et al., 2005). In one case, the mutation was identified by exome sequencing and confirmed by Sanger sequencing. Variants are numbered according to RefSeq transcripts NM_001848.2 for *COL6A1*, NM_001849.3 for *COL6A2*, and NM_004369.3 for *COL6A3*. To examine the diversity of all variants seen in the TH domain of *COL6A1*, *COL6A2*, and *COL6A3*, we compared all coding variants in the TH domain of 417 individuals referred for clinical genetic testing at the University of Utah with annotated variants in dbSNP, build 135 (Sherry, et al., 2001). For variant analyses, familial alleles were considered only once per family. Variants reported here have been submitted to the Leiden Muscular Dystrophy Pages for *COL6A1* (<http://www.LOVD.nl/COL6A1>), *COL6A2* (<http://www.LOVD.nl/COL6A2>), and *COL6A3* (<http://www.LOVD.nl/COL6A3>).

To determine whether the observed clustering of glycine substitutions was predicted by factors in the primary sequence, we developed two different simulations generating glycine substitutions in the TH domain based on the distribution of glycine codons in the primary sequence and on the neighbor-dependent predicted mutation rates for each glycine codon (Hess, et al., 1994). First, to test the hypothesis that the observed N-terminal clustering of glycine substitutions in each of the 3 genes was not predicted by the primary sequence, we simulated glycine substitution in each gene based on the total number of observed glycine substitutions for that gene (93 for *COL6A1*, 33 for *COL6A2*, and 22 for *COL6A3*). In each round of the simulation, the number of glycine substitutions N-terminal to TH17 was counted for each gene. After 50,000 rounds, an empiric distribution of the count of glycine mutations N-terminal to TH17 was determined for each gene and compared to the observed count for that gene. Mean and 95% confidence intervals were determined based on the empiric distribution. Second, to test the hypothesis that the observed clustering of glycine mutations in the $\alpha 1$ (VI) chain vs. the $\alpha 2$ (VI) and $\alpha 3$ (VI) chains is not dependent on the primary sequence we focused our simulation on the region N-terminal to TH17, and simulated 133 glycine substitutions on $\alpha 1$ (VI), $\alpha 2$ (VI), or $\alpha 3$ (VI), counting the number of glycine substitutions per gene for each of 10,000 cycles. We determined a mean and 95% confidence interval for the number of glycine mutations in each gene based on the empiric distribution and compared to the observed count.

Mutations in glycine codons (GGA, GGC, GGG, GGT) can result in substitution of 8 different amino acids or one stop codon. To test whether glycine mutations in our cohort were enriched for any of the possible amino acid substitutions, we calculated expected values for each substitution based on a total of 148 glycine substitutions in the TH domain using neighbor-dependent predicted mutation rates (Hess, et al., 1994; Persikov, et al., 2004). Expected values for each potential substitution and nonsense mutation were compared to the observed glycine substitutions including our cohort and published cases using a χ^2 test under the null hypothesis that observed glycine substitutions occur at expected rates given neighbor-dependent substitution rates and distribution of glycine codons. Deviations from the observed to expected glycine substitutions were correlated with the published destabilization scale for glycine substitutions in the triple helix (Ala<Ser<Cys<Arg<Val<Glu<Asp<Trp) (Persikov, et al., 2004).

Immunofluorescence analysis of collagen VI in muscle and cultured fibroblasts

Muscle biopsy and cultured skin fibroblast samples were obtained from existing clinical samples. Dual labeling of 9 μ m frozen muscle sections was performed for co-localization of collagen VI with collagen IV, a marker for the basement membrane. Collagen VI labeling was performed using a rabbit anti-collagen VI polyclonal antibody (a gift from Dr. Rupert Timpl) at 1:3,000 dilution and Alexa Fluor 488-conjugated goat anti-mouse immunoglobulin (Molecular Probes, Eugene, OR) at 1:500 dilution. Basement membrane labeling was performed using a monoclonal mouse anti-collagen IV (Chemicon, Temecula, CA) at 1:1500 dilution, and Alexa Fluor 568-conjugated goat anti-rabbit (Molecular Probes, Eugene, OR) at 1:500 dilution. Images were obtained using a Leica SP5 confocal microscope. Immunofluorescence analysis of cultured fibroblasts was performed as previously described (Lampe, et al., 2008). Briefly, fibroblasts were grown to 90% confluence and then treated with L-ascorbic acid (50ng/ μ l) for 5 days and then fixed in 4% paraformaldehyde and blocked with 10% fetal bovine serum albumin with or without 0.1% TritonX-100. Staining for collagen VI was performed using anti-collagen VI monoclonal antibody MAB3303 (Chemicon, Temecula, CA) at 1:2,500 dilution, and Alexa Fluor 568-conjugated goat anti-mouse immunoglobulin (Molecular Probes, Eugene, OR) at 1:500 dilution. Images were obtained using a Nikon Eclipse Ti microscope.

RESULTS

Patients and Clinical phenotype

We identified 194 individuals with mutations resulting in glycine substitutions in the TH domain of *COL6A1*, *COL6A2*, and *COL6A3*. Included are 97 newly reported cases from 83 different families and 97 published cases from 65 different families (Fig. 2). Phenotype data sufficient to classify individuals as ES-UCMD, typical UCMD, INT-UCMD, or BM was available for 51 newly reported cases and 52 published cases, for a total of 103 cases (Supp. Tables S1, S2). Clinical characteristics of the newly reported cases are summarized in Table 1. Pace *et al.* have suggested that patients with glycine substitutions in a critical region including Gly-X-Y triplets 10-15 have a more severe disruption of collagen VI assembly and a more severe clinical phenotype (Pace, et al., 2008). In our cohort, patients with mutations inside the critical region tend to a more severe phenotype with 48% ES-UCMD or typical

UCMD compared to 23% ES-UCMD or typical UCMD in cases with mutations outside the critical region. Patients with glycine substitutions outside the critical region tended toward a milder phenotype with 40% BM compared to 7% BM in patients with mutations inside the critical region (Fig. 3, Table 2).

We identified two unaffected carriers that were identified after the mutation was first identified in an affected child. In one case the mutation was demonstrated to be acting recessively in combination with a nonsense mutation (patient 61, Supp. Table S1). In the second case, the unaffected father (patient 29) of a patient with an INT-UCMD phenotype (patient 28) carried the same mutation on *COL6A1* (c.859G>A; p.G287R) that was seen in his son. Peak height ratios from sequence trace data suggest that this mutation may be mosaic in the father (Supp. Figure S1).

Variant analysis

Glycine substitutions in *COL6A1*, *COL6A2*, and *COL6A3* are a common mutation type in our clinical sequencing effort. Of 604 patients sequenced at the University of Utah and Prevention Genetics clinical genetic testing services, 166 have known pathogenic mutations, and 39% (64/166) of these are glycine substitutions in the TH domain. In the 194 cases presented here, glycine substitutions occurred at 44 of the 332 potential sites in the TH domain, 28 of which are newly reported. In 148 unrelated cases (83 from our cohort and 65 from published cases), 89% (131/148) were clustered N-terminal to TH17 (Fig. 4A). We considered whether non-glycine coding variants in the TH domain identified in our clinical sequencing effort or in dbSNP build 135 were also clustered on the N-terminal to TH17. In both cases, non-glycine variants were relatively evenly distributed between the three genes and within the TH domain (Fig. 4B). Ten variants resulting in glycine substitutions in the TH domain are reported in dbSNP 135 including 5 on *COL6A1*, 2 on *COL6A2*, and 3 on *COL6A3* (Supp. Table S3). Five of these variants originate from published pathogenic variants reported in OMIM and also reported here. The variant rs11701912 on *COL6A1* (p.G332S) is pathogenic and dominant, but was reported without allele frequency data or other information on phenotype. The remaining variants were reported in exome sequencing projects and are in the C-terminal end of the TH domain where they are expected to act recessively.

To determine whether the clustering of glycine substitutions in the TH domain is determined by factors in the primary sequence, we simulated glycine substitutions in $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains based on neighbor-dependent mutation rates (Hess, et al., 1994), the distribution of glycine codons in each gene, and the observed number of glycine substitutions in each chain. We performed two simulations, first to address the clustering of substitutions within each chain N-terminal to TH17, and second to address the clustering of substitutions on the $\alpha 1(VI)$ chain vs. the $\alpha 2(VI)$ and $\alpha 3(VI)$ chains. In the first simulation, the observed clustering of the glycine substitutions N-terminal to TH17 was highly significant for each gene, with the observed number of glycine substitutions in this region exceeding even the maximum number seen in 50,000 simulations (Table 3). In the second simulation, the observed glycine substitutions in the $\alpha 1(VI)$ chain were significantly more than the expected (86 observed vs. 44 expected; Table 4). The $\alpha 2(VI)$ chain had

approximately the expected number of glycine substitutions (28 observed vs. 33 expected), whereas the $\alpha 3(\text{VI})$ chain showed a paucity of glycine substitutions (17 observed vs. 54 expected).

We next scrutinized whether any of the 8 potential amino acid substitutions or 1 potential nonsense mutation were occurring more frequently than expected given the sequence context of the glycine codons, and whether the substituted amino acid is associated with severity. In the 148 observed, unrelated cases, there was a significant overabundance of Gly>Arg and Gly>Asp substitutions and a significant underrepresentation of Gly>Ala and Gly>Ser substitutions compared to predictions based on the sequence context (Fig. 5), $\chi^2=70.9$; $p=3.3\times 10^{-12}$. A spectrum of clinical phenotypes was seen for all glycine substitutions (Table 5). While the number of cases is too small to make definite conclusions, the three cases with Ala or Ser substitutions were all relatively mild: 2 presented with BM phenotypes and 1 with INT. These substitutions are the least disruptive to the structure of the TH. It has been suggested in other collagen disorders that these milder substitutions may have a milder phenotype and may even remain unascertained (Persikov, et al., 2004). With the exception of the underrepresented Ala and Ser substitutions there are no obvious differences in severity based on the substituted amino acid or the gene containing the substitution (Table 6).

Clinical variation in patients with identical mutations

The most commonly observed mutation in our cohort is the p.G284R substitution in the $\alpha 1(\text{VI})$ chain due to the c.850G>A mutation in *COL6A1*. This mutation is seen in the context of a CpG, which may provide an explanation for its frequent occurrence. While most individuals with the p.G284R mutation had typical UCMD or INT-UCMD phenotypes, patients with both mild (BM) and severe (ES-UCMD) phenotypes were also seen. The c.850G>C mutation also resulting in p.G284R was seen in two related individuals (twins), both with INT-UCMD phenotypes. Twenty-eight individuals with the p.G284R substitution have clinical information sufficient to classify their phenotype including: 5 ES-UCMD, 9 typical UCMD, 12 INT-UCMD, and 2 BM. The second most common substitution, p.G290R on the $\alpha 1(\text{VI})$ chain, was seen in 18 cases due to two different mutations (16 cases from c.868G>A and 2 cases from c.868G>C) while 2 cases showed the c.869G>A mutation, resulting in a p.G290E substitution. Phenotype data were available in 9 of these cases with 1 ES-UCMD, 2 typical UCMD, and 6 INT-UCMD. Substitutions at p.G281 of the $\alpha 1(\text{VI})$ chain showed the most diversity with three different substitutions in 7 cases (4 p.G281R; 2 p.G281E, and 1 p.G281A).

Inheritance

Glycine substitutions in the TH domain are dominantly acting in 96% (186/194) of cases in the combined cohort. Three cases of recessively acting glycine substitutions have been previously reported, all with UCMD phenotypes (Supp. Table S2). In all three cases, the glycine mutations identified were outside the critical N-terminal end of the TH domain (triplets 47, 65, and 92). In two of these cases, PL6 (Jimenez-Mallebrera, et al., 2006) and PL45 (Brinas, et al., 2010), the glycine substitution was homozygous. In the third case, PL51 (Brinas, et al., 2010), the patient was compound heterozygous for two glycine

substitutions in the $\alpha 2(VI)$ chain, one at the 47th Gly-X-Y triplet, and another at the 77th Gly-X-Y triplet. Family data was not reported. The more N-terminal of these two glycine substitutions (p.G394E, Gly-X-Y triplet 47) is situated outside the critical region and was also seen in patient 49 in our cohort (Supp. Table S1). Although a dominant mechanism is likely, this patient has a second allele on the *COLA3* gene that disrupts in the X-position of the 13th Gly-X-Y triplet in the critical region of *COL6A3*. This variant has been seen in the dbSNP135 (rs113331139; minor allele frequency, 1/4545=0.02%), but pathogenicity has not been documented. Lending support to the notion that a glycine substitution at the 47th Gly-X-Y triplet of the $\alpha 2(VI)$ chain is dominant, a glycine substitution in patient 63 on the 47th triplet of the $\alpha 1(VI)$ chain (c.1184G>T; p.G395V) also appears to be acting in a dominant fashion.

We report two new recessive cases. Patient 61 (Supp. Table S1) has a recessive glycine substitution in the 91st Gly-X-Y triplet of $\alpha 3(VI)$ (c.6931G>A, p.G2311R). The second allele is a nonsense mutation in the Y position of the 8th Gly-X-Y triplet (c.6181C>T, p.R2061X). While both mutations are apparent on sequencing from genomic DNA, sequencing of the fibroblast derived cDNA from this patient does not show the nonsense allele, consistent with nonsense-mediated decay of this allele (data not shown). The nonsense mutation in this case is *de novo*, while the glycine substitution is inherited from the clinically unaffected mother. Patient 77 has a recessive nonsense mutation (c.7066G>T; p.G2356X) in the glycine position of the 106th Gly-X-Y triplet of *COL6A3*. This is the only known nonsense mutation of a glycine in the TH domain. The second allele (c.5044delC; p.Q1682Sfs12X) introduces a frameshift in the N2 domain. Parental DNA was not available for further analysis of these alleles. In both cases the patient had a typical UCMD phenotype.

In addition to the glycine substitution, we identified rare variants not annotated in dbSNP 135 in 10 cases in our cohort (Supp. Table S1). In most cases, these are likely rare polymorphisms. In one case (patient 75, Supp. Table S1), we questioned whether the second variant is acting as a modifier. This patient was compound heterozygous for two variants on *COL6A1*, a glycine substitution (c.1022G>A; p.G341D) and a novel second allele (c.1763C>T; p.P588L). The glycine substitution is dominant and has been previously reported in a family with a limb-girdle muscular dystrophy phenotype (Supp. Table S2, PL12) (Scacheri, et al., 2002). Another substitution (p.G341V) at the same codon has also been reported in patients with BM (Supp. Table S2, PL4 and PL16) (Lampe, et al., 2005; Luciola, et al., 2005). The potential modifying allele results in a substitution of leucine for proline (p.P588L) in the X position in the second to last triplet (Gly-Pro-Pro) in the TH domain. Gly-Pro-Pro triplets in this region of other collagens are thought to be important for nucleation of the triple helix (Hyde, et al., 2006) and a mutation here may be acting as a modifier.

Immunofluorescence of collagen VI in muscle and cultured skin fibroblasts

Muscle biopsies from existing clinical samples were available for our analysis from 10 patients, all with dominantly acting glycine mutations. Findings were typical of other collagen VI myopathy patients including presence of collagen VI immunofluorescence in

the ECM, but loss of co-localization to the basement membrane (Fig. 6A). Collagen VI immunofluorescence in cultured fibroblasts was available for 26 patients. Findings were similar in patients with BM, INT-UCMD, typical UCMD, or ES-UCMD phenotypes and include an overall decrease of collagen VI in the ECM, often with a knot-like appearance, and marked intracellular retention (Fig. 6B).

DISCUSSION

Glycine substitutions in the conserved Gly-X-Y motif in the TH domain of collagen VI are the most commonly identified mutations in the collagen VI myopathies, accounting for almost 1/3 of all pathogenic mutations in *COL6A1*, *COL6A2*, and *COL6A3*. Here, we present the hitherto largest clinical and genetic study of patients with glycine substitutions in the TH domain of collagen VI $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$, including 97 newly identified patients and 97 patients identified from review of the literature. Intermediate phenotypes were the most common presentation, accounting for almost half of patients, emphasizing the importance of intermediate phenotypes to the overall phenotypic spectrum of collagen VI myopathy patients. The p.G284R mutation in the $\alpha 1(VI)$ chain was the single most commonly observed mutation with 28 known cases. While both extremes of the clinical spectrum (ES-UCMD and BM) were seen in patients with this mutation, half had an intermediate phenotype, mirroring the cohort as a whole. The broad range in severity for this and other frequently occurring mutations suggests the possibility for modifying genes as has been recently reported in Duchenne muscular dystrophy (Flanigan, et al., 2012). It is also possible that variants within the genes encoding collagen VI are modifiers of severity of clinical symptoms. In most cases these likely represent rare polymorphisms as we have reported them here (Supp. Tables S1, S2). In at least one case (Supp. Table S1, patient 75), the second allele, a substitution of leucine for proline in the penultimate Gly-X-Y triplet in *COL6A1* is suspicious for acting as a modifier. We did not observe a correlation between clinical severity and which of the three chains was mutated nor were we able to demonstrate a clear correlation of severity with the substituting amino acid, although a significant underrepresentation of Gly>Ser and Gly>Ala mutations was observed.

The majority of glycine substitutions in the TH domain are *de novo* dominantly acting mutations. Rare recessive cases have been reported and were also seen in our cohort. While dominant cases are clustered in the N-terminal end of the TH domain, mutations in recessive cases are near the C-terminal end of the TH domain. It has been suggested that the dominant negative mechanism in collagen VI myopathies reflects the ability of mutant chains to be incorporated into the tetramer subunits (Lampe, et al., 2008). The presence of recessive cases, all with mutations in the C-terminal end of the TH domain, suggests that such positional factors within the TH domain are important in the initial formation of the triple helical monomer and incorporation vs. exclusion of mutant chains in the assembly process. It is not entirely clear whether the recessive glycine substitutions in the C-terminal TH domain are so severely disruptive to the triple helix assembly that they are not incorporated into higher order structures (similar to in frame deletions in the region (Lampe, et al., 2008)) or whether the mutant chains are simply less disruptive and thus better tolerated, although the former seems more likely.

The most striking feature of the glycine substitutions in collagen VI is the N-terminal clustering, with 89% of these mutations N-terminal to the 17th Gly-X-Y triplet (TH17). This important landmark is delineated by cysteine residues in the $\alpha 3$ (VI) chain which form disulfide bonds stabilizing tetramers (Fig. 1). Of the 148 unrelated cases, 89% (131/148) are in this N-terminal segment of the TH domain, accounting for only 15% of the length of the TH domain. N-terminal clustering of glycine substitutions in the TH domain was prominent for all 3 chains, but the $\alpha 1$ (VI) chain showed significantly more glycine substitutions than $\alpha 2$ (VI) or $\alpha 3$ (VI). The three most commonly observed mutations (p.G284R, p.G290R and p.G293R) are all on the $\alpha 1$ (VI) chain and account for 35% of all glycine known substitutions in the TH domain. Mutations at these sites are all associated with a CpG motif; however, factors in the primary sequence such as CpG motifs do not predict the magnitude of the observed excess of mutations on *COL6A1*, nor do they predict the clustering of these mutations on *COL6A2* or *COL6A3*. We considered whether the TH domain N-terminal to TH17 was more prone to variation generally; however, rare variants, including both those reported in dbSNP 135 and those seen in our clinical sequencing effort, do not show clustering N-terminal to TH17. The excess of glycine substitutions in this region generally, and on the $\alpha 1$ (VI) chain specifically may reflect a region of local hypermutability beyond that predicted solely by the local sequence. Alternatively, mutations in this region may be more severely disruptive to the assembly and function of collagen VI and thus more likely to result in a clinical phenotype.

Structural analyses have shown that the C-terminal segments of the TH domain are closely associated with each other in the dimer forming an antiparallel 75nm supercoiled helix, and leaving a 30nm segment of the N-terminal TH that is not overlapped with other chains (Fig 1) (Baldock, et al., 2003; Knupp and Squire, 2001). Within this relatively unencumbered segment, Pace *et al.* proposed a critical region including Gly-X-Y triplets 10-15 that is important in microfibril formation and influences severity of disease (Pace, et al., 2008). Patients with more severe clinical symptoms had glycine mutations in the critical region, secreted non-disulfide bonded collagen VI tetramers, and had a severe compromise of microfibril formation. In contrast, patients with mutations outside this region were more mildly affected, secreted disulfide-bonded collagen VI tetramers, and had less severe compromise of microfibril assembly. In our cohort, 59% (88/148) of glycine substitutions occur in this critical region, accounting for only 5% of the length of the TH domain. Consistent with the assertion that this region is critical to assembly and function of collagen VI, patients with mutations in this region in our cohort tend toward a more severe phenotype. Almost half (33/68) of patients with mutations in this region have severe (ES-UCMD or typical UCMD) phenotypes compared to 22% (8/35) of patients with mutations outside the critical region.

Taken together, our findings suggest that clustering of glycine substitutions in the N-terminal region of collagen VI is not based on features of the primary sequence. We hypothesize that this region may represent a functional domain within the triple helix. It has been proposed that the 30nm segment at the N-terminal end of TH domain containing the critical region forms a loop around the N-terminal globular domain of the adjacent tetramer during microfibril formation (Beecher, et al., 2011). A glycine substitution in this region

may result in mutant tetramers that are unable to form normal associations with adjacent N-terminal globular domains. It remains to be seen whether disruptions of the TH in this region are steric, limiting association of tetramers to form microfibrils, or whether glycine mutations in this region also affect the overall binding associations of collagen VI with other ECM components.

In summary, we report a large cohort of patient with glycine substitutions in the Gly-X-Y motif in the triple helical domain of collagen VI. Glycine substitutions in the TH domain are a common mutational mechanism in collagen VI myopathies and have complex inheritance patterns and molecular pathogenesis. Similar to exon skipping mutations in the TH domain (Lampe, et al., 2008), glycine substitutions cluster N-terminal to the 17th Gly-X-Y triplet and are associated with a spectrum of clinical severity. While not fully predictive, glycine substitutions in the critical region from Gly-X-Y triplets 10-15 tend to result in more severe clinical phenotypes. This dissection of the most prevalent mutational mechanism of collagen VI significantly advances our understanding of the molecular genetic mechanisms underlying these disorders and has obvious implication for the development of molecular treatment approaches, such as the allele specific knockdown approaches recently proposed (Gualandi, et al., 2012). Appropriate anticipatory care and genetic counseling in collagen VI myopathy patients should include an assessment of mode of inheritance and clinical phenotype, which is at least in part predicted by the position of the glycine substitution in the TH domain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We are grateful to the patients and families for their participation in the study. The research was supported by funds from the MDA Clinical Research Training Grant to RJB, Primary Children's Medical Center Foundation to RJB, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, through Grant 5K12HD001410-10 to RJB, Muscular Dystrophy Campaign (ARF), and by intramural funds of NINDS to CGB.

REFERENCES

- Baker NL, Morgelin M, Pace RA, Peat RA, Adams NE, Gardner RJ, Rowland LP, Miller G, De Jonghe P, Ceulemans B. Molecular consequences of dominant Bethlem myopathy collagen VI mutations. *Ann Neurol.* 2007; 62(4):390–405. others. [PubMed: 17886299]
- Baker NL, Morgelin M, Peat R, Goemans N, North KN, Bateman JF, Lamande SR. Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum Mol Genet.* 2005; 14(2):279–93. [PubMed: 15563506]
- Baldock C, Sherratt MJ, Shuttleworth CA, Kielty CM. The supramolecular organization of collagen VI microfibrils. *J Mol Biol.* 2003; 330(2):297–307. [PubMed: 12823969]
- Beecher N, Roseman AM, Jowitt TA, Berry R, Troilo H, Kammerer RA, Shuttleworth CA, Kielty CM, Baldock C. Collagen VI, conformation of A-domain arrays and microfibril architecture. *J Biol Chem.* 2011; 286(46):40266–75. [PubMed: 21908605]
- Brinas L, Richard P, Quijano-Roy S, Gartioux C, Ledeuil C, Lacene E, Makri S, Ferreira A, Maugendre S, Topaloglu H. Early onset collagen VI myopathies: Genetic and clinical correlations. *Ann Neurol.* 2010; 68(4):511–20. others. [PubMed: 20976770]

- Camacho Vanegas O, Bertini E, Zhang RZ, Petrini S, Minosse C, Sabatelli P, Giusti B, Chu ML, Pepe G. Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc Natl Acad Sci U S A*. 2001; 98(13):7516–21. [PubMed: 11381124]
- Chu ML, Pan TC, Conway D, Kuo HJ, Glanville RW, Timpl R, Mann K, Deutzmann R. Sequence analysis of alpha 1(VI) and alpha 2(VI) chains of human type VI collagen reveals internal triplication of globular domains similar to the A domains of von Willebrand factor and two alpha 2(VI) chain variants that differ in the carboxy terminus. *EMBO J*. 1989; 8(7):1939–46. [PubMed: 2551668]
- Chu ML, Pan TC, Conway D, Saitta B, Stokes D, Kuo HJ, Glanville RW, Timpl R, Mann K, Deutzmann R. The structure of type VI collagen. *Ann N Y Acad Sci*. 1990a; 580:55–63. [PubMed: 2337306]
- Chu ML, Zhang RZ, Pan TC, Stokes D, Conway D, Kuo HJ, Glanville R, Mayer U, Mann K, Deutzmann R. Mosaic structure of globular domains in the human type VI collagen alpha 3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J*. 1990b; 9(2):385–93. others. [PubMed: 1689238]
- Clement EM, Feng L, Mein R, Sewry CA, Robb SA, Manzur AY, Mercuri E, Godfrey C, Cullup T, Abbs S. Relative frequency of congenital muscular dystrophy subtypes: Analysis of the UK diagnostic service 2001–2008. *Neuromuscul Disord*. 2012; 22(6):522–7. others. [PubMed: 22480491]
- Dang N, Murrell DF. Mutation analysis and characterization of COL7A1 mutations in dystrophic epidermolysis bullosa. *Exp Dermatol*. 2008; 17(7):553–68. [PubMed: 18558993]
- Flanigan KM, Ceco E, Lamar KM, Kaminoh Y, Dunn DM, Mendell JR, King WM, Pestronk A, Florence JM, Mathews KD. LTBP4 genotype predicts age of ambulatory loss in duchenne muscular dystrophy. *Ann Neurol*. 2012 others.
- Gualandi F, Manzati E, Sabatelli P, Passarelli C, Bovolenta M, Pellegrini C, Perrone D, Squarzone S, Pegoraro E, Bonaldo P. Antisense-induced messenger depletion corrects a COL6A2 dominant mutation in Ullrich myopathy. *Hum Gene Ther*. 2012; 23(12):1313–8. others. [PubMed: 22992134]
- Hess ST, Blake JD, Blake RD. Wide variations in neighbor-dependent substitution rates. *J Mol Biol*. 1994; 236(4):1022–33. [PubMed: 8120884]
- Hicks D, Lampe AK, Barresi R, Charlton R, Fiorillo C, Bonnemann CG, Hudson J, Sutton R, Lochmuller H, Straub V. A refined diagnostic algorithm for Bethlem myopathy. *Neurology*. 2008; 70(14):1192–9. others. [PubMed: 18378883]
- Hyde TJ, Bryan MA, Brodsky B, Baum J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J Biol Chem*. 2006; 281(48):36937–43. [PubMed: 16998200]
- Jimenez-Mallebrera C, Maioli MA, Kim J, Brown SC, Feng L, Lampe AK, Bushby K, Hicks D, Flanigan KM, Bonnemann C. A comparative analysis of collagen VI production in muscle, skin and fibroblasts from 14 Ullrich congenital muscular dystrophy patients with dominant and recessive COL6A mutations. *Neuromuscul Disord*. 2006; 16(9-10):571–82. others. [PubMed: 16935502]
- Jobsis GJ, Keizers H, Vreijling JP, de Visser M, Speer MC, Wolterman RA, Baas F, Bolhuis PA. Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nat Genet*. 1996; 14(1):113–5. [PubMed: 8782832]
- Knupp C, Squire JM. A new twist in the collagen story--the type VI segmented supercoil. *EMBO J*. 2001; 20(3):372–6. [PubMed: 11157744]
- Kuo HJ, Maslen CL, Keene DR, Glanville RW. Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. *J Biol Chem*. 1997; 272(42):26522–9. [PubMed: 9334230]
- Lamande SR, Morgelin M, Selan C, Jobsis GJ, Baas F, Bateman JF. Kinked collagen VI tetramers and reduced microfibril formation as a result of Bethlem myopathy and introduced triple helical glycine mutations. *J Biol Chem*. 2002; 277(3):1949–56. [PubMed: 11707460]
- Lampe AK, Bushby KM. Collagen VI related muscle disorders. *J Med Genet*. 2005; 42(9):673–85. [PubMed: 16141002]

- Lampe AK, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Laval SH, Marie SK, Chu ML, Swoboda K, Muntoni F. Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J Med Genet.* 2005; 42(2):108–20. others. [PubMed: 15689448]
- Lampe AK, Zou Y, Sudano D, O'Brien KK, Hicks D, Laval SH, Charlton R, Jimenez-Mallebrera C, Zhang RZ, Finkel RS. Exon skipping mutations in collagen VI are common and are predictive for severity and inheritance. *Hum Mutat.* 2008; 29(6):809–22. others. [PubMed: 18366090]
- Lucioli S, Giusti B, Mercuri E, Vanegas OC, Lucarini L, Pietroni V, Urtizbarea A, Ben Yaou R, de Visser M, van der Kooi AJ. Detection of common and private mutations in the COL6A1 gene of patients with Bethlem myopathy. *Neurology.* 2005; 64(11):1931–7. others. [PubMed: 15955946]
- Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat.* 2007; 28(3):209–21. others. [PubMed: 17078022]
- Merlini L, Martoni E, Grumati P, Sabatelli P, Squarzoni S, Urciuolo A, Ferlini A, Gualandi F, Bonaldo P. Autosomal recessive myosclerosis myopathy is a collagen VI disorder. *Neurology.* 2008; 71(16):1245–53. [PubMed: 18852439]
- Nadeau A, Kinali M, Main M, Jimenez-Mallebrera C, Aloysius A, Clement E, North B, Manzur AY, Robb SA, Mercuri E. Natural history of Ullrich congenital muscular dystrophy. *Neurology.* 2009; 73(1):25–31. others. [PubMed: 19564581]
- Okada M, Kawahara G, Noguchi S, Sugie K, Murayama K, Nonaka I, Hayashi YK, Nishino I. Primary collagen VI deficiency is the second most common congenital muscular dystrophy in Japan. *Neurology.* 2007; 69(10):1035–42. [PubMed: 17785673]
- Pace RA, Peat RA, Baker NL, Zamurs L, Morgelin M, Irving M, Adams NE, Bateman JF, Mowat D, Smith NJ. Collagen VI glycine mutations: perturbed assembly and a spectrum of clinical severity. *Ann Neurol.* 2008; 64(3):294–303. others. [PubMed: 18825676]
- Persikov AV, Pillitteri RJ, Amin P, Schwarze U, Byers PH, Brodsky B. Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. *Hum Mutat.* 2004; 24(4):330–7. [PubMed: 15365990]
- Pescucci C, Longo I, Bruttini M, Mari F, Renieri A. Type-IV collagen related diseases. *J Nephrol.* 2003; 16(2):314–6. [PubMed: 12768082]
- Reed UC, Ferreira LG, Liu EC, Resende MB, Carvalho MS, Marie SK, Scaff M. Ullrich congenital muscular dystrophy and Bethlem myopathy: clinical and genetic heterogeneity. *Arq Neuropsiquiatr.* 2005; 63(3B):785–90. [PubMed: 16258657]
- Scacheri PC, Gillanders EM, Subramony SH, Vedanarayanan V, Crowe CA, Thakore N, Bingler M, Hoffman EP. Novel mutations in collagen VI genes: expansion of the Bethlem myopathy phenotype. *Neurology.* 2002; 58(4):593–602. [PubMed: 11865138]
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29(1):308–11. [PubMed: 11125122]
- Ullrich O. Kongenitale, atonisch-sklerotische Muskeldysrophie, ein weiterer typus der heredonodegenerativen Erkankungen des neuromuskularen Systems. *Z Gesamte Neurol Psychiat.* 1930; 126:171–201.

Figure 1. Domain structure and assembly of collagen VI

(A) Domain structure of collagen VI $\alpha 1$ (VI), $\alpha 2$ (VI), and $\alpha 3$ (VI) chains. N-terminal von Willebrand factor A domains are light gray and C-terminal von Willebrand factor A domains are dark gray. Cysteine residues in the 30th Gly-X-Y triplet of *COL6A1* and *COL6A2* and 17th triplet of *COL6A3* important for higher order assembly are labeled 'S'. The 17th Gly-X-Y triplet (delineated with a dashed line) is an important landmark with 89% of glycine substitutions N-terminal this site. A cysteine residue at this site in the $\alpha 3$ (VI) chain in this is important for disulfide binding of tetramers. (B) Assembly of collagen VI highlighting, first, association of $\alpha 1$ (VI), $\alpha 2$ (VI), and $\alpha 3$ (VI) chains to form the monomer subunit, second, antiparallel association of monomers stabilized by disulfide bonds between cysteine residues in the 30th Gly-X-Y triplet of $\alpha 1$ (VI) and $\alpha 2$ (VI) chains with the C-terminal globular domains of the adjacent monomer, and third, parallel association of dimers stabilized by disulfide bonds formed between cysteine residues in the 17th triplet of adjacent $\alpha 3$ (VI) chains. Final assembly into beaded microfibrils occurs by extracellular association of tetramers in end-to-end fashion (Chu, et al., 1990b; Pace, et al., 2008). Size of the globular N- and C-terminal domains and the length of the TH domain are proportional to size estimations proposed by Beecher et al.(Beecher, et al., 2011)

Figure 2. Summary of known cases with glycine substitutions in the TH domain of $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains (194 cases)

Triangles above the axis represent glycine substitutions identified in our cohort and below the axis represent glycine substitutions identified in published cases. Related individuals are linked by a bracket. Position of cysteine residues important for dimerization ($\alpha 1(VI)$, $\alpha 2(VI)$), and tetramerization ($\alpha 3(VI)$) are marked with a red box. The critical region important for assembly proposed by Pace et al. is marked with a horizontal orange box. *indicates both alleles in one patient with two different glycine substitutions. + represents patients homozygous glycine substitutions. # indicates alleles known/suspected to be acting in recessive fashion. TH domain exons are indicated by alternating dark and light grey rectangles.

Figure 3. Clinical phenotype and position of mutation for patients with glycine substitutions in the TH domain (103 cases)

Triangles above the axis represent glycine substitutions identified in our cohort and below the axis represent glycine substitutions identified in published cases. Related individuals are linked by a bracket. Position of cysteine residues important for dimerization ($\alpha1(VI)$, $\alpha2(VI)$), and tetramerization ($\alpha3(VI)$) are marked with a red box. The critical region important for assembly proposed by Pace et al. is marked with a horizontal orange box. + represent patients homozygous glycine substitutions. * denotes both alleles in one patient with two different glycine substitutions, # indicates alleles known or suspected to be acting in recessive fashion. TH domain exons are indicated by alternating dark and light grey rectangles.

Figure 4. Distribution of glycine and non-glycine substitutions on the $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chain

Position of cysteine residues important for dimerization ($\alpha 1(VI)$, $\alpha 2(VI)$), and tetramerization ($\alpha 3(VI)$) are marked with a small vertical box. The critical region important for assembly proposed by Pace et al. is marked with a horizontal box. (Pace, et al., 2008)

(A) Clustering of glycine substitutions in TH domain of $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains. Tic marks above the axis represent alleles identified in our cohort. Tic marks below the axis represent alleles identified in published cases. The position of CpG motifs with potential to cause a glycine substitution are marked with diamonds. The three most commonly observed substitutions (p.G284R, p.G290R, and p.G293R) are all in the *COL6A1* gene and are all in the context of CpG motifs. Glycine substitutions in the context of CpG elsewhere in the TH domain were only seen at 2 of the possible 43 CpG sites (*COL6A2*: c.1450G>A, p.G484R; *COL6A3*: c.6175G>T p.G2059C). Neither of these has more than 1 independent observation among the 148 unrelated cases. (B) Distribution of intermediate frequency variants (0.05%-5% minor allele frequency, full height tic), and low frequency variants (<0.05% minor allele frequency, half height tic) in the TH domain of $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains. Non-glycine missense variants observed in the TH domain in our clinical sequencing effort (tic marks above the axis) and in dbSNP135 (tic marks below the axis) are not clustered on *COL6A1*, nor are they clustered within the TH domain.

Figure 5. Distribution of observed and expected glycine substitutions in $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains in 148 unrelated cases based on neighbor dependent mutation rates
Order of substituted amino acids on the X-axis is based on the increasing severity of disruption to the stability of the triple helix.

Figure 6. Immunofluorescent staining of muscle (A) and cultured fibroblasts (B) from a patient with c.812G>A; p.G271D mutation in *COL6A2*

Dual labeling of collagen VI (green) and collagen IV (red) in muscle shows presence of collagen VI in the ECM, but poor co-localization with the basement membrane (yellow). Scale bar is 75 μ m. Inset shows higher magnification for patient muscle (A,b). Staining for collagen VI in cultured fibroblasts from the same patient shows decreased deposition and speckled appearance of collagen VI in the ECM of patient fibroblasts compared to control (B;a,b). Collagen VI staining in the presence of TritonX-100 to permeabilize the cell membrane demonstrates intracellular retention of collagen VI in patient fibroblasts (B: c,d).

Clinical summary of 51 patients with glycine substitutions in the TH domain of $\alpha 1(VD)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ **Table 1**

Clinical Phenotype	Number	Age ^a	CK ^b	Age walked	Avg. first wheelchair (yr)	Avg. full wheelchair (yr)	FTT ^c	Respiratory compromise ^d
ES-UCMD	2 (4%)	4.5 (3-6)	214	-	-	-	1 (50%)	1 (50%)
UCMD	10 (20%)	11.7 (5-19)	245 (116-395)	19.8	7.8	9.2	4 (40%)	8 (80%)
INT	24 (44%)	11.6 (4-41)	302 (193-470)	16.8	10.8	15.0	11 (46%)	5 (21%)
BM	11 (20%)	28.0 (8-50)	1000 (500-1500)	14.8	-	-	0	1 (9%)
RECESSIVE	2 (2%)	7.5 (5-10)	-	18.0	-	10.0	2 (100%)	1 (50%)
UNAF ^e	2 (4%)	49.5 (38-61)	-	-	--	-	-	-

^a Age at last clinical evaluation^b CK = mean and range of creatine kinase in patients from which it was available^c FTT = failure to thrive, defined as weight under 5th centile for age^d Forced expiratory volume (FEV1) <60% predicted for age, or non-invasive ventilation, or tracheotomy^e Clinically unaffected patient with glycine substitution

Table 2

Correlation of severity of clinical phenotype by position of glycine substitutions within or outside of critical region (Gly-X-Y triplets 10-15)

Substitution in critical region ^a	ES-UCMD	UCMD	INT	BM	UNAF ^b	Total
No		6	12	14	1	33
Yes	9	24	29	5	1	68
	9	30	41	19	2	101

^a Critical region as defined by Pace et al. from Gly-X-Y triplets 10-15(Pace, et al., 2008)

^b Clinically unaffected patient with glycine substitution

Table 3

Observed and simulated glycine substitutions on $\alpha 1(VI)$, $\alpha 2(VI)$, OR $\alpha 3(VI)$ chains based on neighbor-dependent mutation rates demonstrating clustering of mutants in region N-terminal to TH17 in all 3 chains.

	Observed glycine substitutions in TH domain	Observed glycine substitutions N-terminal to TH17	Simulated substitutions N-terminal to TH17 ^a		
			Avg +/- 2SD ^b	Max	Min
$\alpha 1(VI)$	93	86	17.5 (10-25)	34	3
$\alpha 2(VI)$	33	28	4.2 (0.3-8)	13	0
$\alpha 3(VI)$	22	17	4.6 (0.8-8.4)	13	0

^c Max and min indicate the maximum and minimum counts for the number of substitutions N-terminal to TH17 for each chain from all 50,000 cycles.

^a 50,000 cycles per gene, each cycle placing the observed number of glycine substitutions from the entire TH domain and counting the number of substitutions in the region N-terminal to TH17.

^b Average count of glycine substitutions N-terminal to TH17 in 50,000 simulations for each gene. 95% confidence interval in parenthesis.

Table 4

Observed and simulated glycine substitutions in the region N-terminal to TH17 for $\alpha 1(VI)$, $\alpha 2(VI)$, and $\alpha 3(VI)$ chains based on neighbor-dependent mutation rates demonstrating clustering of substitutions on $\alpha 1(VI)$.

Observed substitutions		Simulated substitutions^a		
		Avg $\pm 2SD$^b	Max^c	Min
$\alpha 1(VI)$	86	43.9 (33.1-54.7)	69	23
$\alpha 2(VI)$	28	33.0 (23.1-42.9)	59	13
$\alpha 3(VI)$	17	54.1 (42.8-65.4)	77	32

^a 10,000 cycles with each cycle placing 131 glycine substitutions in the region N-terminal to TH17 on any of the three chains with count of the number of substitutions for each chain in each cycle.

^b Average count of glycine substitutions in each chain in 10,000 simulations. 95% confidence interval in parenthesis.

^c Max and min indicate the maximum and minimum counts for each chain from the 10,000 cycles.

Table 5

Severity of clinical phenotype by the substituted amino acid.

Substitution ^a	ES-UCMD	UCMD	INT	BM	UNAF ^b	Total
A			1	1		2
S				1		1
C	2	2		2		6
R	7	19	28	10	2	66
V		2	2	3		7
E		3	4			7
D		4	7	2		13
W						
X		1				1
Total	9	31	42	19	2	103

^a Amino acid substitutions are arranged in increasing disruption to triple helix.(Persikov, et al., 2004)^b Clinically unaffected patient with glycine substitution

Table 6

Severity of the clinical phenotype by gene.

Gene	ES-UCMD	UCMD	INT	BM	UNAF ^a	Total
<i>COL6A1</i>	6	17	30	10	1	64
<i>COL6A2</i>	1	9	6	2	-	18
<i>COL6A3</i>	2	6	5	7	1	21
Total	9	32	41	19	2	103

^a Clinically unaffected patient with glycine substitution