The amino acid sequence of several human globin chains [37] was determined in the late 1950s and early 1960s by direct protein sequencing prior to the advent of gene cloning and DNA sequencing. In these original sequences, the first amino acid of the human α-, β- and δ-globins is valine and that of γ-globin is glycine. However, HGVS nomenclature numbers the amino acids beginning with the methionine encoded by the initiation codon. Consequently, the sickle-cell disease β-globin variant, in which glutamic acid is replaced by valine, should be reported as being at position 7, rather than 6, according to HGVS recommendations. Indeed, this variant is still described in OMIM (Online Mendelian Inheritance in Man) [38] and in the HbVar database for hemoglobin variants thalassemia mutations [39,40] in terms of the legacy amino acid numbering scheme.

Even though non-standard, the legacy numbering of the globin amino acids is well recognized by experts in the field. However, this is not true for newcomers or students who may blindly assume that standards are being applied and may become either completely lost or waste valuable time sorting out the problem. The same is also true in the case of phosphoglycerate kinase 1 (PGK, encoded by the PGK1 gene), where considerable confusion has arisen from describing variants in relation to alterations to the known mature amino acid sequence [41]. Again, the issue arises because PGK is one of the few enzymes in which variants were characterized at the amino acid level prior to DNA sequencing being widely used.

The collagens also provide excellent examples of legacy numbering schemes. Because of the lure of the characteristic triple-helical nature of the collagens, numbering of the amino acids was established decades ago with the first glycine of the (Gly-X-Y)n-repeat region being designated as amino acid number 1. In addition, when the first genomic DNA clones were isolated, exons were initially numbered in the 3′ to 5′ direction, a lack of full-length cDNA clones hampering the determination of the exact number of exons. Consequently, the first osteogenesis imperfecta variant that was characterized was reported as being in exon 1 of the COL1A2 gene, which encodes the α2 chain of type I collagen [42]. In fact, the gene is now known to comprise 52 exons and the variant lies in exon 52 using conventional numbering. However, other exon-numbering anomalies remain. The COL1A1 and COL1A2 genes that encode the alpha chains of type I collagen are evolutionarily related but COL1A1 has a single exon that corresponds to exons 33 and 34 of COL1A2. This single exon is known as exon 33/34 [43] and the designation, which is more than 20 years old, is still widely used in the current literature.

A further issue is the discovery of additional exons in genes where an exon-numbering scheme has already been established. This has resulted in the opioid receptor, mu 1 gene (OPRM) having exons designated O, X and Y, with exons 3 and 5 divided into two and five sub-regions, respectively [44], and the cystic fibrosis transmembrane conductance regulator gene (CFTR) having exons designated 6a, 6b, 14a, 14b, 17a and 17b [45].