Murphy's Law and the Human Beta-Globin Gene

Charles L. Vigue

Charles L. Vigue is an associate professor and chair of the Dept. of Biology at the Univ. of New Haven, West Haven CT 06516. He received his B.A. and M.S. in Zoology from the Univ. of Maine and his Ph.D. in Genetics from North Carolina State Univ. He held an N.I.H. postdoctoral fellowship at The Johns Hopkins Univ. He has published in several journals, including *Genetical Research, Biochemical Genetics*, the *American Biology Teacher* and the *Journal of College Science Teaching*. He is an active member of the Connecticut Science Teachers Assoc. and a science scripter for the Connecticut Radio Information Service, a radio station broadcasting the printed media to the handicapped.

Human alpha-globin and beta-globin, two of the best characterized human genes, can be used to illustrate virtually all the concepts of normal eukaryotic gene structure and function. As with most other genetic systems, the understanding of normal gene structure and function has been made possible by the analysis of mutations. Murphy's Law, "Whatever Can Go Wrong Probably Will," certainly applies to the human globin genes. Mutations affecting virtually every aspect of globin synthesis and function have been found. This paper reviews normal beta-globin structure and function and some of the many mutations that affect it. Alpha-globin mutations are used when there are no corresponding beta-globin examples.

Normal Human Beta-Globin Gene

Human beta-globin is a protein of 146 amino acids encoded by a gene on chromosome 11. Two beta-globin chains and two alpha-globin chains comprise the human hemoglobin molecule. Like other eukaryotic proteins, beta-globin is encoded by a DNA sequence (Figures 1 and 2) which is divided into sequences appearing in the mature mRNA (exons) and sequences not appearing in the mature mRNA (intervening sequences, IVSs, or introns).

The most common sequence of the beta-globin gene is that given by Lawn, Efstratiadis, O'Connell & Maniatis (1980) (Figure 2). This sequence is designated as framework 1. Several other normal frameworks exist as polymorphisms, however (Antonarakis, Kazazian & Orkin 1985). The variant polymorphisms and distributions in various ethnic groups are given in Tables 1 and 2, respectively.

To the 5′ side of the start of transcription (nucleotide 1) at positions −26 to −34, the promoter or so-called "TATA box" is located. (Bases to the 5′ side of the start site of transcription are designated as "minus.") The promoter consists of the sequence:

5′ GGCATAAAA 3′
−34 −26

and is thought to be responsible for the binding of RNA polymerase. Other sequences 5′ to the promoter such as the "CCAAT box" located from −72 to −76 also influence transcription but in an unknown manner (see below).

The RNA polymerase transcribes a pre-mRNA

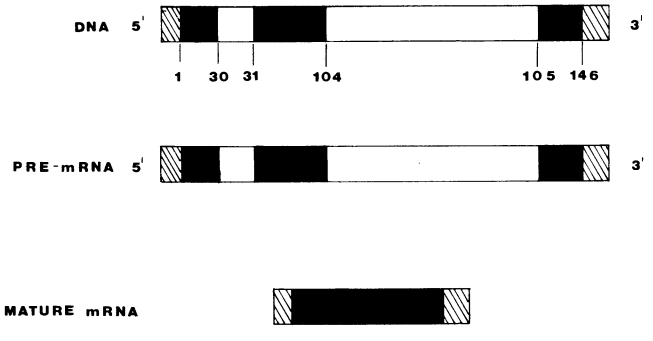

Figure 1. The Human Beta-Globin Gene, Its Pre-mRNA and Mature mRNA. Striped areas represent transcribed but not translated regions. Open areas represent intervening sequences (IVS-1 is 5′ to IVS-2.). Black areas represent the exons (Exon 1 is the more 5′, and exon 3 is more 3′.). Numbers represent the beta-globin amino acid residues encoded by the three exons.

Table 1. Normal Polymorphism of the Human Beta Globin Gene

Framework	*Polymorphism*
2	G to T at IVS-2 position 74
3	G to T at IVS-2 position 74
	C to G at IVS-2 position 16
	C to T at IVS-2 position 81
	T to C at IVS-2 position 666
	C to T at third nucleotide of codon 2, exon 1
3 (Asian)	G to T at IVS-2 position 74
	C to G at IVS-2 position 16
	T to C at IVS-2 position 666
	C to T at third nucleotide of codon 2, exon 1

Figure 2. The Nucleotide Sequence of the Sense Strand of Beta-Globin DNA. Amino acids corresponding to the sense codons are indicated above the codon. Nucleotide 1 (the beginning of transcription) and codon 1 are underscored and indicated by number. The promoter, initiator codon, translation terminator codon and the transcription termination signal are underscored. (Adapted from Lawn et al., 1980).

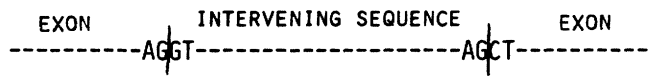

Figure 3. Splice Site Junctions. Solid vertical lines represent exon/intervening sequence boundaries.

(also called heterogeneous nuclear (hn) RNA) (Figure 1) which includes a 5′ leader sequence of about 50 nucleotides, the 3 exons of 90, 122 and 126 bases, two introns of 130 and 850 bases and a 3′ trailer sequence of about 130 bases. An AATAAA sequence located 17 to 25 bases from the 3′ end of the pre-mRNA signals the termination of transcription.

The main event after transcription is the removal (splicing out) of the intervening sequences to give a mature mRNA (Figure 1). Splicing normally takes place at specific AG/GT and AG/CT boundaries as indicated in Figure 3.

Missense Mutations

Mutations resulting in amino acid substitutions are called missense mutations. The first missense mutation to be identified for any protein (gene) was the beta-globin mutation which causes sickle cell anemia. In sickle cell anemia, codon number 6 (exon 1) of the beta-globin gene has mutated from GAG to GTG resulting in the substitution of valine for the normal glutamic acid. Sickle cell is the most common hemoglobin varient in the United States. More than 200 missense mutations have been identified in beta-globin (Nora & Fraser 1981). A partial list and their effects are given in Table 3.

Nonsense Mutations

Nonsense mutations are those mutations that change codons specifying amino acids (sense codons) to termination codons (nonsense codons). Hemoglobin McKees Rocks results from codon 145 (TAT, exon 3), which normally codes for tyrosine, mutating to TAA, a nonsense codon (Nora & Fraser 1981). Consequently the last two amino acids (numbers 145 and 146) are missing from the completed beta-globin chain. Mild anemia results.

Table 2. Distribution of the Different Frameworks in Various Ethnic Groups

	Framework			
Group	*1*	*2*	*3*	*3 (Asian)*
Mediterranean	53%	18%	19%	
American Blacks	79	12		9%
Asiatic Indians	52	16		32
South East Asians	18	35		47
Chinese	21	27		52

Severe anemia such as beta-thalassemia can result, however, if the nonsense mutation occurs earlier (to the 5′ end) of the hemoglobin gene. For example, beta-thalassemia has been shown to result from an AAG (lysine) to TAG (termination) mutation in codon 17 (exon 1) (Chang & Kan 1979), a CAG (glutamine) to TAG (termination) mutation in codon 39 (exon 2) (Moschonas, de Boer, Grosveld & Dahl 1981) and a TGG (tryptophan) to TGA (termination) mutation in codon 15 (exon 1) (Kazazian, Chakravarti & Orkin 1983). The resulting beta-globin fragments are nonfunctional.

Initiation Codon Mutations

The initiation codon, ATG, which codes for formylmethionine could conceivably mutate to another codon, and thereby reduce or eliminate transcription. No initiation codon mutations have been found for beta-globin, but an ATG to ACG (thr) initiation codon mutation has been identified for alpha-globin. This mutation results in alpha-thalassemia (Pirastu, Sagho, Cao & Kan 1984).

Termination Codon (Chain Elongation) Mutations

Theoretically, the TAA termination codon at the end of the coding portion of the beta-globin mRNA could mutate to a sense codon resulting in an extension of the protein chain by continuous reading of the mRNA. To date, no termination codon mutations have been found for beta-globin. An alpha-globin

Table 3. Some Missense Mutations in Human Beta Globin

Name of Disease	*Number*	*Normal Amino Acid*	*Mutant Amino Acid*	*Effect*
Hemoglobin Koln	98	Val	Met	Instability of Globin Molecule
Hemoglobin Hammersmith	42	Phe	Ser	Instability of Globin Molecule
Hemoglobin Bristol	67	Val	Asp	Instability of Globin Molecule
Hemoglobin C	6	Glu	Lys	Instability of Globin Molecule
Hemoglobin Kansas	102	Asn	Thr	Decreased Affinity for O_2
Hemoglobin Seattle	70	Ala	Asp	Decreased Affinity for O_2
Hemoglobin Olympia	20	Val	Met	Increased Affinity for O_2
Hemoglobin Athens, GA	28	Leu	Pro	Increased Affinity for O_2
Hemoglobin Saskatoon	63	His	Tyr	Inability of Fe to Remain Reduced

Table 4. Some Beta-Globin Gene Deletions

Name of Disorder	*Residue Number(s)*	*Amino Acid(s)*
Hemoglobin Leiden	6 or 7	Glu
Hemoglobin Lyon	17 and 18	Lys and Val
Hemoglobin Freiburg	23	Val
Hemoglobin Niteroi	42-44(43-45)	Phe-Gly-Ser
Hemoglobin Tochigi	56-59	Gly-Asn-Pro-Lys
Hemoglobin Tours	87	Thr
Hemoglobin Gun Hill	91-95(92-96; 93-97)	Cys-Asp-Lys, Leu,Lys
Hemoglobin Leslie	131	Gln

termination codon mutation, hemoglobin Constant Spring, however, results in the alpha-globin having 31 extra amino acids.

Deletion Mutations

Several base deletion mutations in the beta-globin gene resulting in amino acid deletions are known. Table 4 summarizes some of the known deletions. The deletions often cause severe distortions in the hemoglobin molecule resulting in instability. Effects range from mild to severe anemia.

The deletions are thought to arise by mispairing of the homologous chromosomes followed by unequal crossing-over as indicated in Figure 4 for hemoglobin Freiburg.

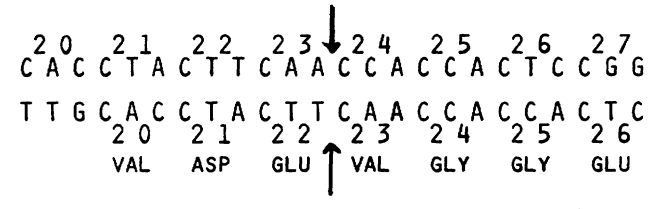

Figure 4. Unequal Crossing Over and the Generation of Hemoglobin Freiburg. Numbers refer to codon numbers. Arrows show points of crossing over.

Promoter Mutations

Two mutations in the promoter region that result in beta-thalassemia have been found. One is an A to G transition at position −29 (Antonarakis et al. 1984), and the other is an A to C transversion at position −28 (Poncz et al. 1982). Both mutations reduce the level of transcription to only 25 percent of normal.

Mutations 5′ to Promoter

Treisman, Orkin & Maniatis (1983) found a C to G transversion at position −87 that reduces transcription to about 10 percent of normal levels. The effects of this mutation clearly indicates that sequences 5′ to the promoter influence gene function. The exact nature of these sequences, however, is unknown.

Frameshift Mutations

One cause of beta-thalassemia is the insertion of an adenine residue between codons 71 and 72 (Cheng et al. 1984). This causes a shift in the reading frame and the creation of new codons resulting in an in phase termination codon at position 73 as indicated in Figure 5. No normal hemoglobin is synthesized.

```
CODON NUMBER                    70  71  72  73
NORMAL BASE SEQUENCE          - GCC TTT AGT GAT ---
NORMAL AMINO ACID SEQUENCE    - ALA PHE SER ASP ---
FRAMESHIFT BASE SEQUENCE      - GCC TTT AAG TGA ---
FRAMESHIFT AMINO ACID SEQUENCE - ALA PHE LYS
```

Figure 5. Frameshift Mutation Caused by an Insertion of an A Residue between Codons 71 and 72. The inserted A is underlined.

Hemoglobin Tak is a frameshift mutation resulting from the insertion of an AC dinucleotide between codons 146 [CAC(His)] and 147 [TAA(ter)] (Figure 6) (Weatherall & Clegg 1986). This frameshift extends the carboxy terminus of the hemoglobin molecule by eleven amino acids. An in phase termination codon is in position 158. Anemia results from the unstable, elongated hemoglobin. Frameshift mutations more 5′ than hemoglobin Tak would probably result in beta-thalassemia.

Mutations Inactivating Splice Sites

An A to G transition at the next to last (#849) base of IVS-2 inactivates the highly conserved AG splice site sequence so that removal of the intervening sequence does not occur (Figure 7) (Antonarakis et al 1984). No normal hemoglobin is synthesized.

A G-to-A transition at the first position of IVS-2 similarly inactivates the GT splice site (Orkin et al. 1982). IVS-2 is not excised. Beta-thalassemia results since no normal hemoglobin is synthesized.

Mutations Creating Alternate Splice Sites

A G-to-A mutation at position 1 of IVS-1 eliminates the normal splice site and activates one or more cryptic sites. For example, codon 17 exon 1, which is not normally used as a splice site although it contains an AG dinucleotide, is now used as a 5′ splice site resulting in the excision of 40 percent of exon 1 (codons 18 through 30) along with IVS-1 (Figure 8) (Treisman, Orkin & Manlatis 1983). No normal hemoglobin is synthesized.

Hemoglobin E is a missense mutation in codon 26 [GAG(Glu)] to [AAG(Lys)]. The mutation results in mild anemia due to a reduction in the amount of normal beta-globin mRNA (Antonarakis et al. 1982).

CODON NUMBER	145	146	147	148	149	150	151	152	153	154	155	156	157	158
NORMAL NUCLEOTIDE SEQUENCE	-TAT	CAC	TAA	GCT	CGC	TTT	CTT	GCT	GTC	CAA	TTT	CTA	TTA	AAG
NORMAL AMINO ACID SEQUENCE	-TYR	HIS												
TAK NUCLEOTIDE SEQUENCE	-TAT	CAC	ACT	AAG	CTC	GCT	TTC	TTG	CTG	TCC	AAT	TTC	TAT	TAA
TAK AMINO ACID SEQUENCE	-TYR	HIS	THR	LYS	LEU	ALA	PHE	LEU	LEU	SER	ASN	PHE	TYR	

Figure 6. Frameshift Mutation Resulting in Hemoglobin Tak. The inserted dinucleotide is underlined.

The mutant slowly excised IVS-1 and occasionally alternately splices at a cryptic 3′ site within codon 26 which contains an AG dinucleotide resulting in the excision of codons 27 through 30 (Orkin et al. 1982).

A GGT (Gly) to GGA (Gly) silent mutation in codon 24 also creates an alternate AG dinucleotide splice site since codon 25 is GGT. Beta-thalassemia results from the removal of codons 25 through 30 along with IVS-1 (Goldsmith et al 1983).

The creation of alternate splice sites in intervening sequences could result in the incomplete removal of the intron. A G-to-A mutation 21 nucleotides from the 3′ end of IVS-1 creates an AG dinucleotide. This alternate splice site is preferred to the normal site and results in an additional 19 nucleotides from IVS-1 in the processed mRNA between exons 1 and 2. The extra nucleotides cause a shifting of the reading frame so no normal beta-globin is synthesized from the abnormal mRNA (Figure 9) (Busslinger Moschonas & Flarell 1981).

Transcription Termination Signal Mutations

Orkin et al. (1985) have identified a T-to-A mutation in the AATAAA transcription termination signal. RNA polymerase reads through the normal 3′ termination site creating a pre-mRNA 900 nucleotides longer than normal. Beta-thalessemia results although some translation of the elongated mRNA may occur.

It is obvious from the above discussion of some of the many hemoglobin mutations that the human beta-globin gene obeys Murphy's Law. It is hoped that teachers of general biology, genetics, biochemistry and cell and molecular biology will be able to use these examples of beta-globin mutations in their classes to illustrate normal and abnormal gene function.

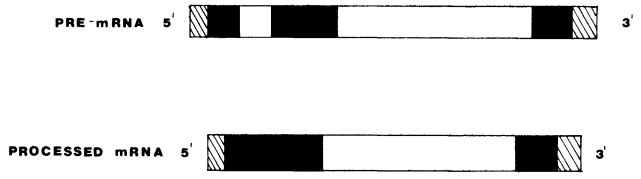

Figure 7. Abnormal Processing of Human Beta-Globin mRNA due to a Splice Site Mutation. Striped, open and solid areas represent untranslated regions, intervening sequences and exons, respectively.

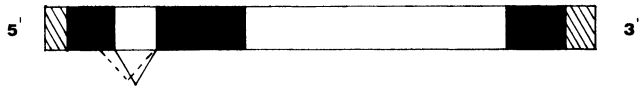

Figure 8. Abnormal Processing of Human Beta-Globin mRNA due to Creation of an Alternate Splice Site within an Exon. Striped, open and solid areas represent untranslated regions, intervening sequences and exons, respectively. The solid line represents normal excision of IVS-1 and the dotted line represents that area that is abnormally excised.

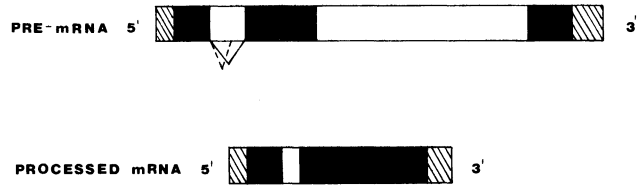

Figure 9. Abnormal Processing of Human Beta-Globin mRNA due to Creation of an Alternate Splice Site within an Intron. The legend is the same as Figure 8.

References

Antonarakis, S., Orkin, S., Kazazian, H., Goff, S., Boehm, C., Waber, P., Sexton, J., Ostrer, H., Fairbanks, V. & Chakravarti, A. (1982). Evidence for multiple origins of the B^E-globin gene in Southeast Asia. *Proceedings of the National Academy of Sciences, 79,* 6608-6611.

Antonarakis, S., Irkin, S., Cheng, T-C., Scott, A., Sexton, J., Trusko, S., Charache, S. & Kazazian, H. (1984). B-Thalassemia in American Blacks; Novel mutations in the "TATA" box and an acceptor splice site. *Proceedings of the National Academy of Sciences, 81,* 1154-1158.

Antonarakis, S., Kazazian, H., & Orkin, S. (1985). DNA polymorphisms and molecular pathology of the human globin gene clusters. *Human Genetics, 69,* 1-14.

Busslinger, M., Moschonas, N. & Flavell, R. (1981). B^+thalassemia: Aberrant splicing results from a single point in an intron. *Cell, 27,* 289-298.

Chang, J. & Kan, Y. (1979). B^o thalassemia, a nonsense mutation in man. *Proceedings of the National Academy of Sciences, 76,* 2886-2889.

Cheng, T-C., Orkin, S., Antonarakis, S., Potter, M., Sexton, J., Markham, A., Giardina, P., Li., A & Kazazian, H. (1984). B-Thalassemia in Chinese: Use of *in vivo* RNA analysis and oligonucleotide hybridization in systemic characterization of molecular defects. *Proceedings of the National Academy of Sciences, 81,* 2821-2825.

Goldsmith, M., Humphries, R., Ley, T., Cline, A., Kantor, J. & Nienhaus, A. (1983). "Silent" nucleotide substitution in a B^+-thalassemia globin gene activates splice sites in coding sequence RNA. *Proceedings of the National*

Academy of Sciences, 80, 2318-2322.
Kazazian, H., Chakravarti, A., & Orkin, S. (1983). DNA polymorphisms in the human B-globin gene cluster. In M. Nei & R. Koehn (Eds.), *Evolution of Genes and Proteins,* (pp. 137-146), Sunderland: Sinauer Associates.
Lawn, R., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980). The nucleotide sequence of the human B-globin gene. *Cell, 21,* 647-651.
Moschonas, N., de Boer, E., Grosveld, F. & Dahl, H. (1981). Structure and expression of a cloned B-thalassemia globin gene. *Nuc. Acid Res., 9,* 4391-4401.
Nora, J. & Fraser, F. (1981). *Medical Genetics.* Philadelphia: Lea and Febiger.
Orkin, S., Cheng, T-C., Antonarakis, S., & Kazazian, H. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of human B-globin gene. *EMBO, 4,* 453.
Orkin, S., Kazazian, H., Antonarakis, S., Goff, S., Boehm, C., Sexton, J., Waber, P. & Giardina, P. (1982). Linkage of B-thalassemia mutations and B-globin gene polymorphisms in human B-globin gene cluster. *Nature, 296,* 627-631.
Orkin, S., Kazazian, H., Antonarakis, S., Ostrer, H., Goff, S. & Sexton, J. (1982). Abnormal RNA processing due to the exon mutations of B^E-globin gene. *Nature, 300,* 768-769.
Pirastu, M., Sagho, G., Cao, A., & Kan, Y.W. (1984). An initiation codon mutation in alpha thalassemia. *Clinical Research, 32,* 550.
Poncz, M., Ballantine, M., Solowiejczyk, D., Barak, I., Schwartz, E. & Surrey, S. (1982). B-Thalassemia in a Kurdish Jew. *Journal of Biological Chemistry, 257,* 5994-5996.
Treisman, R., Orkin, S. & Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned B-thalassemia genes. *Nature, 302,* 591-596.
Weatherall, D. & Clegg, J. (1976). Molecular genetics of human hemoglobin. *Annual Review of Genetics, 10,* 157-178.

Acknowledgement

The author wishes to acknowledge the assistance of Jerry Sangiamo in the preparation of the figures.