Blood Cells, Molecules and Diseases xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

Review

Molecular basis of β thalassemia and potential therapeutic targets

Swee Lay Thein*

Sickle Cell Branch, National Heart, Lung, and Blood Institute/National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

Keywords:
β thalassemia
Genotype/phenotype correlation
Fetal hemoglobin
Genetic prediction

ABSTRACT

The remarkable phenotypic diversity of β thalassemia that range from severe anemia and transfusion-dependency, to a clinically asymptomatic state exemplifies how a spectrum of disease severity can be generated in single gene disorders.

While the genetic basis for β thalassemia, and how severity of the anemia could be modified at different levels of its pathophysiology have been well documented, therapy remains largely supportive with bone marrow transplant being the only cure. Identification of the genetic variants modifying fetal hemoglobin (HbF) production in combination with α globin genotype provide some prediction of disease severity for β thalassemia but generation of a personalized genetic risk score to inform prognosis and guide management requires a larger panel of genetic modifiers yet to be discovered.

Nonetheless, genetic studies have been successful in characterizing the key variants and pathways involved in HbF regulation, providing new therapeutic targets for HbF reactivation. BCL11A has been established as a quantitative repressor, and progress has been made in manipulating its expression using genomic and gene-editing approaches for therapeutic benefits. Recent discoveries and understanding in the mechanisms associated with ineffective and abnormal erythropoiesis have also provided additional therapeutic targets, a couple of which are currently being tested in clinical trials.

1. Introduction

The inherited disorders of hemoglobin (Hb) production are the most common human monogenic disorders, among which those affecting the adult β globin gene (*HBB*) – β thalassemia and sickle cell disease (SCD) – are the most clinically significant [1,2]. β thalassemia is caused by a spectrum of mutations that results in a quantitative reduction of β globin chains that are structurally normal [3], in contrast to SCD which is caused by an abnormal Hb variant (HbS, βGlu6Val) that results from a point mutation in the *HBB* gene [4,5]. This change predisposes HbS to polymerization when deoxygenated, a primary event indispensable in the molecular pathogenesis of SCD.

β thalassemia occurs widely in a broad belt including the Mediterranean, parts of North and sub-Saharan Africa, the Middle East, Indian subcontinent and Southeast Asia. It appears that heterozygotes for β thalassemia [6] are protected from the severe effects of falciparum malaria, and natural selection has increased and maintained their gene frequencies in these malarious tropical and sub-tropical regions. In these prevalent regions, gene frequencies for β thalassemia range between 2 and 30% [6]. However, owing to population movements in recent years, β thalassemia is no longer confined to these high-incidence regions, but have become an important public health problem in many countries, including North America and Europe [1]. Because carriers for α-thalassemia and HbS are similarly protected from the severe effects of falciparum malaria [7,8], regions where β-thalassemia is prevalent overlaps substantially with those of α-thalassemia and HbS. Hence, it is not unusual to encounter individuals who have co-inherited two or more Hb variants; for example, 30–35% of SCD individuals from West Africa have co-inherited α-thalssemia [9], and co-inheritance of α- with β-thalassemia is fairly common in the Mediterranean and South-East Asia [10,11].

2. The β globin gene (*HBB*) and normal expression

β globin is encoded by a structural gene found in a cluster with the other β-like genes on chromosome 11 (11p 15.15) [12]. The cluster contains five functional genes, ε (*HBE*), Gγ (*HBG2*), Aγ (*HBG1*), δ (*HBD*), and β (*HBB*), which are arranged along the chromosome in the order of their developmental expression to produce different Hb tetramers: embryonic (Hb Gower-1 ($\zeta_2\varepsilon_2$), Hb Gower-2 ($\alpha_2\varepsilon_2$), and Hb Portland ($\zeta_2\beta_2$)), fetal ($\alpha_2\gamma_2$), and adult (HbA, $\alpha_2\beta_2$ and HbA_2, $\alpha_2\delta_2$) [12] and Serjeant & Vichinsky 2017, this issue. Expression of the globin genes is dependent on local promoter sequences as well as the upstream β globin locus control region (β-LCR) which consists of five Dnase 1

* Sickle Cell Branch, National Heart, Lung, and Blood Institute, The National Institutes of Health, 10 Center Drive, Building 10, Room 6S241A, Bethesda, MD 20892-1589, USA.
E-mail address: sl.thein@nih.gov.

http://dx.doi.org/10.1016/j.bcmd.2017.06.001
Received 17 May 2017; Received in revised form 9 June 2017; Accepted 9 June 2017
1079-9796/ Published by Elsevier Inc.

hypersensitive (HS) sites (designated HS1 to HS5) distributed between 6 and 20 kb 5′ of *HBE* gene [13–15]. There is one HS site at approximately 20 kb downstream of *HBB* gene. All these regulatory regions bind a number of key erythroid-specific transcription factors, notably GATA-1, GATA-2, NF-E2, KLF1 (also known as EKLF), and SCL as well as various co-factors (e.g. FOG, p300), and factors that are more ubiquitous in their tissue distribution, such as Sp1 [12,16,17] (See Philipsen and Hardison 2017, this issue).

The β-like globin genes are each expressed at distinct stages of development through a process referred to as hemoglobin switching (embryonic → fetal → adult). At six months after birth, fetal Hb (HbF, $\alpha_2\gamma_2$) which comprises < 5% of the total Hb, continues to fall reaching the adult level of < 1% at two years of age, when adult Hb becomes the major Hb, and mutations affecting the adult *HBB* gene, i.e. β thalassemia and SCD become manifested [18–20]. In contrast severe mutations which result in complete absence of α globin genes, become clinically apparent at the fetal stage.

The *HBB* locus is a paradigm for tissue- and developmental stage-specific regulation; expression of the individual globin genes relies on a timely and direct physical interactions between the globin promoters and the β-LCR, the interaction being mediated through binding of erythroid-specific and ubiquitous transcription factors [21,22]. A dual mechanism has been proposed for the developmental expression: 1) gene competition for the upstream β-LCR, conferring advantage for the gene closest to the LCR [23], and 2) autonomous silencing (transcriptional repression) of the preceding gene [24,25]. The ability to compete for the β-LCR and autonomous silencing depends on the change in the abundance and repertoire of various transcription factors that favour promoter-LCR interaction. While the ε and γ globin genes are autonomously silenced at the appropriate developmental stage, expression of the adult β globin gene depends on lack of competition from the upstream γ gene for the LCR sequences. Concordant with this mechanism, when the γ gene is upregulated by point mutations in their promoter causing a non-deletion hereditary persistence of fetal hemoglobin (HPFH), expression of the *cis* β gene is downregulated [26]. Further, mutations which affect the β globin promoter, which removes competition for the β-LCR, are associated with higher than expected increases in γ (HbF, $\alpha_2\gamma_2$) and δ (HbA_2, $\alpha_2\delta_2$) expression [27–29]. In recent years, an increased understanding of the repressors and co-repressors of the γ globin gene (e.g. BCL11A) and the switch from fetal to adult Hb expression has provided much insight on strategies of de-repressing expression of the fetal globin genes in adults for treating both β thalassemia and SCD.

3. Genetics of β thalassemia

> 300 β thalassemia alleles have now been described (http://www.ithanet.eu/db/ithagenes; http://globin.bx.psu.edu/hbvar) but only about forty account for 90% or more of the β thalassemias worldwide [30]. This is because in the areas where β thalassemia is prevalent, only a few mutations are common, possibly reflecting local selection due to malaria. Each of these populations, thus has its own spectrum of β thalassemia alleles.

Downregulation of the β globin gene can be caused by a whole spectrum of molecular lesions ranging from point changes to small deletions limited to *HBB*, to extensive deletions of the whole β globin cluster (Fig. 1) [3]. In contrast to α thalassemia, which is mainly caused by deletions (Harteveld and Farashi 2017, this issue), the vast majority of mutations causing β thalassemia are non-deletional.

Functionally, β thalassemia alleles are considered as β^0 when no β globin is produced, or β^+ in which some β globin is produced, but less than normal. A range of severity is encountered within the β^+ thalassemia group; the less severe forms are sometimes designated β^{++} to reflect the minimal deficit in β chain production. Some β^{++} alleles are so mild that they are 'silent', and carriers do not display any evident hematological phenotypes; their red cell indices and HbA_2 levels are within normal limits, the only abnormality being an imbalanced α:non-α chain synthesis [31]. These β^{++} thalassemia alleles have usually been uncovered in individuals with thalassemia intermedia who have inherited a silent β thalassemia allele in compound heterozygosity with a severe allele. In this case, one parent has typical β thalassemia trait, and the other is apparently normal. In contrast, carriers for β^0 and β^+ thalassemia alleles have clearly recognizable hematological phenotypes – mild or no anemia, microcytic hypochromic red cell indices, and elevated HbA2 with mildly increased HbF levels (Fig. 2).

3.1. Non-deletion β thalassemia

These non-deletional mutations, i.e. single base substitutions, small insertions or deletions of one to a few bases are located within the gene or its immediate flanking sequences. They downregulate the β globin gene via almost every known stage of gene expression, from transcription to RNA processing and translation of β globin mRNA. Approximately half of the non-deletional mutations completely inactivate the β gene with no β globin production resulting in β^0 thalassemia. (See Table 1 for categories of non-deletional mutations.)

3.1.1. Transcriptional mutations

Transcriptional mutants involve the conserved DNA sequences that form the β globin promoter (from 100 bp upstream to the site of the initiation of transcription, including the functionally important CACCC, CCAAT and ATAA boxes) or the stretch of 50 nucleotides in the 5'UTR. Generally, these transcriptional mutants result in a mild to minimal reduction of β globin output i.e. β^+ or β^{++} thalassemia alleles, and occasionally they are 'silent'. A silent β thalassemia allele which has been observed fairly frequently in the Mediterranean region is the − 101C → T mutation where it interacts with a variety of more severe β thalassemia mutations to produce milder forms of β thalassemia [32]. Other 'silent' mutations include those in the 5′ UTR; the extremely mild phenotype is exemplified in a homozygote for the + 1 A → C mutation who has the hematologic values of a thalassemia carrier, heterozygotes are 'silent' [33].

Within this group of transcriptional mutants, ethnic variation in phenotype has been observed. Black individuals homozygous for the − 29 A → G mutation have an extremely mild disease [34], while a Chinese individual homozygous for the same mutation had severe anemia and was transfusion-dependent [35]. The cause of this striking difference in phenotype is not known but likely to be related to the different chromosomal backgrounds on which the apparently identical mutations have arisen. One difference is the C-T polymorphism at position − 158 upstream of the $^{G}\gamma$ globin gene (*Xmn*1-$^{G}\gamma$ site) present in the β chromosome carrying the − 29 A → G mutation in Blacks but absent in that of the Chinese. The *Xmn*I-Gγ site, considered to be a quantitative trait locus for HbF, is associated with increased HbF production under conditions of erythropoietic stress (see later on 'Update on the genetic control on HbF).

3.1.2. Mutations affecting RNA processing

A wide variety of mutations interfere with processing of the primary mRNA transcript. Those that affect the invariant dinucleotide GT or AG sequences at exon-intron splice junctions prevent normal splicing altogether, causing β^0 thalassemia. Mutations involving the consensus sequences adjacent to the GT or AG dinucleotides allow normal splicing to varying degrees and produce a β thalassemia phenotype that ranges from mild to severe. For example, mutations at position 5 IVS1 G → C, T or A, considerably reduce splicing at the mutated donor site compared with the normal β allele [26]. On the other hand, the substitution of C for T in the adjacent nucleotide, intron 1 position 6, only mildly affects normal RNA splicing. Although the IVS1-6 T-C mutation is generally associated with milder β thalassemia, studies have shown differential severities for apparently identical mutations; again this is presumably related to the chromosomal background on which the mutations have