

Dysregulation of β -Globin Gene Expression in β -Thalassaemia

Sarah Lawrence, Kathryn Mahina and Emma Holstein

CXA300 Human Molecular Biology
School of Human Life Sciences

Mentor: Mr Dale Kunde

Abstract

β -thalassaemia is the most common inherited blood disorder, affecting synthesis of the β -globin chain of haemoglobin leading to dysregulation of haemoglobin synthesis and serious pathophysiological syndromes. There are over 200 mutations in the β -globin gene cluster that are known to cause β -thalassaemia. A number of these are mutations that directly affect the regulation of the β -globin gene such as deletions of controlling elements like the locus control region (LCR) or inhibition of regulators that are essential for gene activation and transcription. The focus of this review is the structure, function and regulation of the β -globin gene cluster and genetic defects that can result in dysregulation of the LCR and the control of haemoglobin switching that can lead to β -thalassaemia.

Introduction

The three main forms of haemoglobin present in adult erythrocytes are Hb A, Hb F and Hb A₂. Hb A accounts for the majority of all haemoglobin and normally consists of two α - and two β -globin chains surrounding a haeme moiety. Production of α and β globin chains is highly regulated with equal proportions synthesised for normal erythropoiesis. β -thalassaemia is the most common inherited blood disorder, affecting synthesis of the β -globin chain of haemoglobin.

The pathophysiology of the thalassaemias is almost entirely due to the imbalance of globin chain synthesis (Higgs 2004). The genes encoding the α - and β -globin chains are located on chromosome 16 and chromosome 11 respectively. Patients with β -thalassaemia have mutations in the gene responsible for the production of the β -globin, resulting in no or significantly reduced β -chain synthesis. The result is a lack of haemoglobin produced in the erythroid precursors, and therefore small, hypochromic red blood cells are present. However, more importantly, the excess α -globin chains precipitate in developing red cells, causing damage to surface membranes leading to ineffective erythropoiesis and premature haemolysis of peripheral red blood cells as they are removed by macrophages in the reticulo-endothelial system (Lewis 2001).

The β -Globin Gene Cluster

Structure and function

The β -globin chain is a 146 amino-acid polypeptide, encoded by a structural gene found in a cluster on the short arm of chromosome 11 (Olivieri 1999) over approximately 60 000 bp (Cao 2002). The gene cluster consists of six functional genes arranged in the same order in which they are expressed during development; that is 5'- ϵ -, ζ -, γ -, δ -, β - 3' (Rodak 2002).

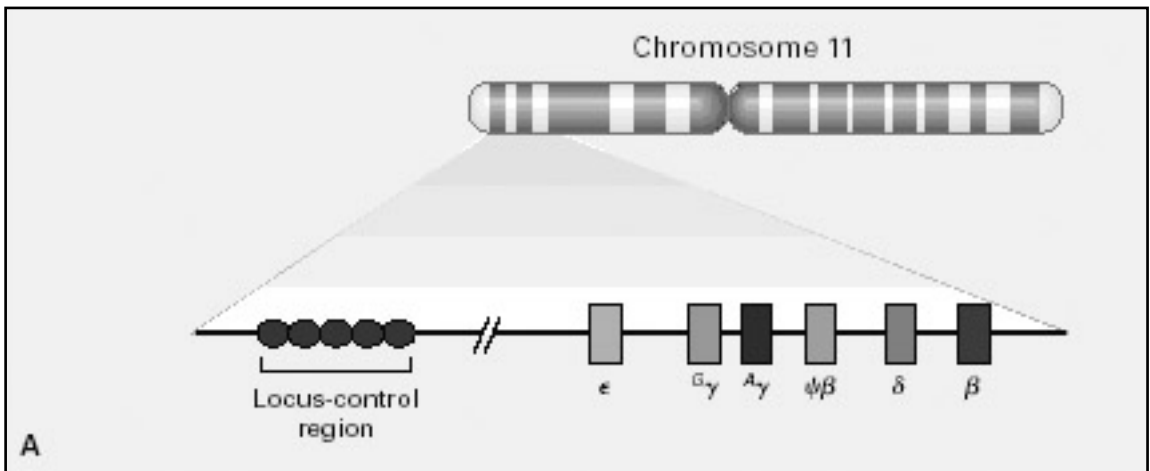


Figure 1: β -globin gene locus (Olivieri 1999). The locus is located on chromosome 11, and consists of a cluster of six globin genes. Upstream of the cluster is the main regulatory element the Locus Control Region.

The transcribed region of the β -globin gene contains three exons separated by two intervening sequences. Exon 2 encodes the residues involved in haeme binding and $\alpha\beta$ dimer formation, while exons 1 and 3 encode for the non-haeme-binding regions of the β -globin chain (Olivieri 1999). Promoter elements upstream of the initiation codon of each active gene are involved in the initiation of transcription. The cluster also contains other regulatory elements that interact to promote erythroid-specific gene expression and to co-ordinate the developmental regulation of each gene (Cao 2002)

The β -globin gene, as well as the other genes in the β -globin cluster, is transcribed in a centromere-to-telomere direction (Jia 2003). DNA sequences required for accurate initiation of the transcription of the β -globin gene are located in the promoter within approximately 200 bp upstream from the transcription start site. The complete sequence of approximately 2.0 kb on chromosome 11 contains most, but not all, of the sequence elements required to encode and regulate the expression of this gene. Many of the important structural features of the β -globin gene, including conserved promoter sequence elements, intron and exon boundaries, RNA splice sites, the initiator and termination codons, and the polyadenylation signal are known to be mutated in various inherited defects of the β -globin gene (Nussbaum 2004).

Expression and regulation

The expression of genes in the β -globin gene cluster is generally regulated in such a way that at any point in development the output of the β -globin-like chains equals that of one of the α -globin chains. This is the result of fine transcriptional, post-transcriptional, and post-translational control. However, these regulatory mechanisms do not take into account whether the genes, which are yet to be activated, are functional. Therefore when the programmed developmental time for the activation of a globin gene is reached, the transcription machinery will be positioned and engaged on its promoter even when the gene is defective and unable to produce any protein. When the defect resides in the β -globin gene, the decreased β -globin synthesis will result in unbalanced chain production and β -thalassaemia (Rodak 2002)

The β -globin gene promoter includes three positive cis-acting elements: a TATA box, a CCAAT box and duplicated CACCC motifs. While the CCAAT and TATA elements are found in many eukaryotic promoters, the CACCC sequence is found predominately in erythroid cell-specific promoters (Dempsey 2003). A transcription factor known as the Erythroid Kruppel-Like Factor (EKLF) binds

to the CACCC motif and it appears to be crucial for normal adult β -globin gene expression. In addition to these motifs, elements upstream of the β -globin promoter contain two binding motifs for the erythroid transcription factor GATA-1. The importance of these 5'-flanking sequences for normal gene expression is underscored by β -thalassaemia arising from point mutations in these sequences; specifically in and around the TATA box and the CACCC motifs in the -80 to -100 region. An enhancer is also found in intron 2 and 3' of the globin gene, 600-900 bp downstream of the poly (A) site (Olivieri 1999)

The 5'-UTR occupies a region of 50 nucleotides between the CAP site, the start of transcription, and the initiation (ATG) codon. There are two prominently conserved sequences in the 5'-UTR of the various globin genes. One is the CTTCTG hexanucleotide found 8-13 nucleotides downstream from the CAP site. The second conserved sequence is CACCATG, in which the last three nucleotides form the initiation codon. Again, the importance of these sequences in the regulation of the β -globin gene expression is exemplified by several mutations in the 5'-UTR causing β -thalassaemia.

The Locus Control Region

Structure

The LCR is the most distal regulatory element in the human β -globin locus. It is comprised of 5 DNase I hypersensitive sites and spans at least 17 kilobases of DNA (Elnitski 2000). There are flanking sequences on either side of each hypersensitive site that help stabilise the complex formed by the LCR during the initiation of transcription. It is a site of DNA-protein interaction and each hypersensitive site contains combinations of several DNA motifs for interacting transcription factors. Of these the most important are GATA-1 and NF-E2 as they confer a high level of position independent expression on linked globin genes in transgenic mice (Cao 2002). HS 1-4 are erythroid specific and have binding sequences for GATA-1 and NF-E2, while HS 5 is ubiquitous (Thein 2004).

Function

The LCR plays a key role in the regulation of β -globin gene expression. It may be involved in enhancing globin gene transcription and insulating the locus from the influence of flanking elements (Tahara 2003 and Thein 2004). HS 5 acts as a chromatin insulator, while the enhancer activity of the LCR resides in HS 2, 3 and 4 (Li 2002). The exact role of HS 1 is yet to be determined. The LCR initiates transcription in an endogenous retroviral long-term repeat that is located upstream of HS 5 (Routledge 2002). There are additional transcription initiation sites in HS 2 and 3. The enhancer function of these hypersensitive sites comes from a 200-300 base pair core within them that house a number of binding sites for ubiquitous and erythroid specific transcription factors (Li 2002). In particular the core of HS 2 is a powerful enhancer of globin gene expression.

Tandem NF-E2 binding sites in this core region are essential for HS 2's enhancer functions and chromatin remodelling activity (Routledge 2002). Routledge and Proudfoot (2002) showed that where alterations to the NF-E2 binding site resulted in a significant decrease in β -globin gene expression and Navas et al (2002) demonstrated that deletions of the flanking sequences of HS 2 and 3 resulted in a modest decrease in β -globin expression. From this it was theorised that the LCR contains functionally redundant elements so if deletions occur at those sites the function of the entire LCR is not adversely affected as the remaining hypersensitive sites will be able to compensate for the deletions (Navas 2002). Deletion of the core elements in either HS 2 or 3 resulted in a marked decrease in β -globin gene expression, which reiterates the importance of these sites for maximal β -globin gene function.

Developmental expression of the individual globin genes is reliant on gene silencing and gene competition, which is controlled through interaction between globin gene promoters and the LCR (Mahajan 2002). Expression of the adult β -globin gene may be dependant on the low level of competition from the γ -globin gene for the LCR sequences, which is autonomously silenced during development. This is supported by the down regulation of the β -globin gene when the γ -globin gene is up regulated by promoter mutations such as those seen in the non-deletional hereditary persistence of foetal haemoglobin (Thein 2004)

Many theories have been postulated regarding the physical mechanisms by which the LCR regulates β -globin gene activity. Four models of LCR function have been proposed: looping, linking, tracking and facilitated tracking.

Looping

The looping model proposes that the individual hypersensitive sites work together to form a large holocomplex (Tang 2002). Within this holocomplex the hypersensitive sites' core elements form active sites that bind transcription factors. Flanking sequences hold the holocomplex in the proper conformation and help to stabilise the structure (Li 2002)

Linking

The linking model suggests that the LCR promotes the formation of a chain of transcriptional facilitator proteins that extends along the chromatin fibre between the LCR and the globin gene family to activate individual genes promoters. This results in the formation of a number of chromatin elements that initially define the area to be transcribed and enhance gene activation (Tang 2002)

Tracking

The tracking model proposes that an activation complex tracks linearly along the DNA until it reaches the basal transcription machinery located at the correct promoter. The complete transcription apparatus is then assembled and transcription of that gene is initiated (Li 2002)

Facilitated Tracking

The facilitated tracking model proposes that a transcription factor and co-activator complex, bound to the LCR, form a loop to contact downstream DNA in regions distal to a promoter. The complex is subsequently released in this region and tracks along the chromatin until it encounters the appropriate promoter (Li 2002). A stable loop structure is then formed and the gene is transcribed.

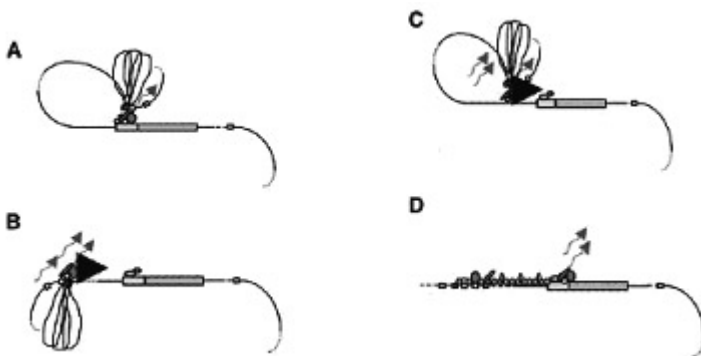


Figure 2: Models of LCR Function (Li 2002). (A) Looping. The holocomplex loops out DNA sequences between the LCR and a genes promoter, delivering transcription factors directly to it. (B) Tracking. An activation complex tracks along the DNA until it reaches a promoter and assembles a transcription apparatus, initiating transcription. (C) Facilitated tracking. A transcription factor complex loops to contact DNA distal to a promoter. It is then released in this region and tracks along the chromatin until it encounters a promoter and transcription is initiated. (D) Linking. The LCR forms a chain of transcriptional facilitator proteins that contact a promoter and initiate transcription.

Mutations That Affect LCR Function

Mutations affecting the LCR can have a dramatic impact on the level of expression of the genes in the β -globin gene cluster. Studies in an Indian population have identified a single nucleotide polymorphism in the binding sites of the core fragments of HS 2, 3 and 4 that may have a causative role in both the β -thalassaemia major and carrier phenotypes (Kukreti 2002). Deletions that remove the entire LCR silence the whole gene cluster and result in changes to chromatin structure that subsequently cause $\epsilon\gamma\delta\beta$ -thalassaemia. Recent studies have also shown that deletions of the individual LCR hypersensitive sites in YAC transgenes results in a significant reduction in globin gene expression. A reduction in HS 1 formation in the β -globin gene promoter along with the reduction of HS 3 formation in the LCR lead to decreased adult β -globin gene expression (Levings 2002). Mutations that remove competition for the LCR, such as those affecting the β -globin gene promoter, are associated with increases in the expression of the γ and δ globin genes (Thein 2004). Mutations that affect the production of transcription factors working in conjunction with the LCR to promote globin gene expression will result in low-level transcription of the β -globin gene cluster.

Haemoglobin Switching

β -thalassaemia can be caused by mutations that affect the control of haemoglobin switching. Haemoglobin switching involves the suppression of the γ -globin gene, which is accompanied by a subsequent rise in the expression of the previously silent β -globin gene (Cao 2002). Haemoglobin switching is regulated primarily by competition between the γ - and β -globin genes' promoters and the LCR. This involves complex interactions between promoter elements and a number of novel transcription factors that have been collectively called switching factors.

Switching Factors

Switching factors are able to regulate the stage-specific expression of γ - and β -globin genes, by binding to the corresponding γ - and β -promoters. Several switching factors have been identified.

The switching factors that are involved in the regulation of embryonic genes are Foetal Kruppel-Like Factors, FKLf and FKLf-2. FKLf's are preferentially expressed in foetal erythroid tissue (Gabbianelli 2000) and appears to be a transactivator for the ϵ -globin genes, enhancing ϵ -globin gene expression to six times the normal rate. FKLf-2 is the transactivator for the γ -globin gene and can enhance its expression up to 40 times that of its inactivated level. The expression profiles of FKLf and FKLf-2 are not absolutely erythroid specific and their role in switching needs further confirmation in experiments on gene disruption (Cao 2002).

EKLF is an erythroid-specific activator that is critical for switching on high-level adult β -globin expression. EKLF binds to the CACCC element in the proximal β -promoter and leads to chromatin remodelling and transcriptional activation (Zhao 2004). EKLF is expressed more in the adult stages of development (Anderson 2000) and its gene inactivation in mice leads to a clear picture of β -thalassaemia (Cao 2002). It also has been shown that mutations in the CACCC element resulting from thalassaemia, interfere with EKLF binding and activation (Marini 2004). This may result in decreased β -globin gene expression.

The stage selector protein complex (SSP), a protein that binds to the stage selector element (SSE) of the γ -globin promoter, may act as a haemoglobin switching factor. It has been shown to confer foetal stage expression on a γ -globin promoter and to delay and prolong haemoglobin switching when deleted from transgenic mouse constructs (Cao 2002).

Nuclear Factor E4 (NF-E4), is another globin-specific transcription factor that appears to confer preferential activation of the γ - over the β -globin promoter (Cao 2002). Enforced expression of NF-E4 leads to increased foetal globin gene expression. Where active competition between foetal

and adult globin genes occurs, enforced expression of NF-E4 leads to a reduction in β -globin gene expression (Zhao 2004).

There are two other factors that may act as haemoglobin switching factors, as they have been shown to influence γ -globin gene expression. COUP-TFII, is a retinoic acid orphan receptor that may act as a repressor of the γ -globin expression. The other PYR, is an adult stage-specific factor that binds to a pyrimidine-rich DNA stretch between the γ - and β -globin genes from where it might repress γ -globin and activate β -globin gene expression. Deletion of the PYR element causes prolonged and delayed haemoglobin switching (Cao 2002).

Reactivation of γ -globin genes

The deletions that cause β -thalassaemia can differ greatly in the size of the deletion, however the majority of these involve the removal of a region in the 5'- β promoter, which includes the CACCC, CCAAT and TATA elements. These deletions in heterozygous β -thalassaemias appear to increase the expression of γ and δ genes by the elimination of competition between the upstream LCR and limiting transcription factors. This leads to increased levels of Hb A₂ and Hb F due to increased γ - and δ -chains being available for haemoglobin synthesis. The increased γ and δ gene expression is variable, which results in variable increases in Hb F levels. However in the case of β -thalassaemia homozygotes, the elevated Hb F levels can partially compensate for the absence of β -globin and decreased Hb A levels. This mechanism may also explain the unusually high HbA₂ levels that accompany point mutations affecting the β promoter region (Thein 2004).

Mechanisms of Globin Gene Reactivation

In one study conducted (Li 2004), several observations were made which involve Hb switching and possible mechanisms of regulation. One conclusion was that the placing of a γ -globin gene between the LCR and the β -globin gene results in β -globin gene expression in embryonic erythroid cells being reduced or absent. This supports the gene competition model of Hb switching.

When the LCR enhancer is in proximity of a gene, the equilibrium between the LCR enhancing activity and the promoter silencing activity will favour gene activation. Therefore, any globin genes placed in close proximity of the LCR, will be expressed during any developmental stage, regardless of the developmental specificity. High-level γ -globin gene expression will continue during adult erythropoiesis when the γ -globin gene is in the proximity of the LCR. The presence of even a minimal promoter is therefore able to induce high levels of gene expression (Li 2004).

When the distance between the γ -globin gene and the LCR exceeds a certain point, the gene is either no longer expressed or there is minimal amount of gene expression on adult erythroid cells. The main developmental regulation mechanism of the β -globin gene appears to be transcription suppression, which occurs by modulating the competency of the promoters for interaction with the LCR (Li 2004).

It was proposed that modulation of Hb F may result from signal transduction involving fully committed erythroblasts (Bhanu 2004). This hypothesis arose from the demonstration that Hb F and Hb A, which are both elevated in β -thalassaemia, share a coordinated rather than a "switched" expression pattern during human adult erythropoiesis. The study showed that certain stem cell factors (CD117) had a variable affect on Hb F.

Therapeutic reactivation of γ -globin genes

Pharmacologic reactivation of γ -globin expression may provide an effective treatment for patients with β -thalassaemia. In thalassaemia, a significant increase in γ -globin production could partially replace the defective β -globin, thus preventing precipitation of unpaired α -globin chains (Gabbianelli 2000). There are several recent studies (Gabbianelli 2000 and Thein 2004) focused



on isolation and identification of compounds that would have the ability to reactivate the γ -globin gene expression. It is assumed that the major effects of such drugs will be on transcription rates, although effects on mRNA processing, stability or other post-transcriptional mechanisms cannot be excluded. At the present, no single compound has been shown to fulfil these requirements without other clinical situations arising, such as cytotoxicity (Smith 2000). However the future is full of potential and other possible inducers or stimulators of haemoglobin reactivation (Gabbianelli 2000) are being investigated for their applications to clinical situations.

References:

Anderson, K.P., Crable, S.C., Lingrel, J.B. (2000) The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression, *Blood*, 95:1652-1655

Bhanu, N.V., Trice, T.A., Lee, Y.J., Miller, J.L. (2004) A signaling mechanism for growth-related expression of fetal hemoglobin, *Blood*, 103(5):1929-1933

Cao, A., Moi, P. (2002) Regulation of the Globin Genes, *Pediatric Research*, 51(4):415-421

Dempsey, N.J., Ojalvo, L.S., Wu, D.W., Little, J.A. (2003) Induction of an embryonic globin gene promoter by short-chain fatty acids, *Blood*, 102(12):4214-4222

Elnitski, L., Li, J., Noguchi, C.T., Miller, W., Hardison, R. (2000) A Negative Cis-element Regulates the Level of Enhancement by Hypersensitive Site 2 of the β -Globin Locus Control Region, *The Journal of Biological Chemistry*, 276(9):6289-6298

Gabbianelli, M., Testa, U., Massa, A., et al. (2000) Hemoglobin switching in unicellular erythroid culture of sibling erythroid burst-forming units: kit ligand induces a dose-dependent fetal hemoglobin reactivation potentiated by sodium butyrate, *Blood*, 95(11):3555-3561

Higgs, D.R. (2004) Gene Regulation in Hematopoiesis: New Lessons from Thalassemia, *Hematology 2004: American Society of Hematology*, 1:1-13

Jia, C.P., Huang, S.Z., Yan, J.B., Xiao, Y.P., Ren, Z.R., Zeng, Y.T., (2003) Effects of human locus control region elements HS2 and HS3 on human β -globin gene expression in transgenic mouse, *Blood Cells, Molecules, and Diseases*, 31:360–369

Kukreti, R., B-Rao, C., Das, S.K., et al (2002) Study of the Single Nucleotide Polymorphism (SNP) at the Palindromic Sequence of Hypersensitive Site (HS) 4 of the Human β -Globin Locus Control Region (LCR) in Indian Population, *America Journal of Haematology*, 69:77-79.

Levings, P.P., Bungert, J. (2002) The human β -globin locus control region: A center of attraction, *European Journal of Biochemistry*, 269:1589-1599

Lewis, S.M., Bates, I., Bian, B.J. Dacie, J.V. (2001) Dacie and Lewis: Practical Haematology, 9th Ed, Churchill Livingstone, London, UK. p. 236

Li, Q., Peterson, K.R., Fang, X., Stamatoyannopoulos, G. (2002) Locus control regions, *Blood*, 100(9):3077-3086

Li, Q., Han, H., Ye, X., Stafford, M., Barkess, G., Stamatoyannopoulos, G. (2004) Transcriptional potentials of the β -like globin genes at different developmental stages in transgenic mice and hemoglobin switching, *Blood Cells, Molecules, and Diseases*, 33:318– 325

- Mahajan, M.C., Weissman, S.M. (2002) DNA-dependent adenosine triphosphatase (helicase-like transcription factor) activates β -globin transcription in K562 cells, *Blood*, 99(1):348-356
- Marini, M.G., Asunis, I., Porcu, L., et al. (2004) The distal β -globin CACCC box is required for maximal stimulation of the β -globin gene by EKLF, *British Journal of Haematology*, 127:114–117
- Navas, P.A., Li, Q., Peterson, K.R., et al (2002) Activation of the β -like globin genes in transgenic mice is dependent on the presence of the β -locus control region, *Human Molecular Genetics*, 11(8):893–903
- Nussbaum, R.L., McInnes, R.R., Willard, H.F. (2004) *Thompson and Thomson: Genetics in Medicine*, 6th Ed, Saunders, Pennsylvania, USA. p. 193-201
- Olivieri, N.F. (1999) The β -Thalasseмииs, *Medical Progress*, 341(2):99-109
- Rodak, B.F. (2002), *Haematology, Clinical Principles and Applications*, 2nd edition, W.B. Saunders Company, USA, p. 343-348
- Routledge, S.J.E., Proudfoot, N.J., (2002) Definition of Transcriptional Promoters in the Human β Globin Locus Control Region, *Journal of Molecular Biology*, 323:601–611
- Tahara, T., Sun, J., Nakanishi, K., et al. (2003) Heme Positively Regulates the Expression of β -Globin at the Locus Control Region via the Transcriptional Factor Bach1 in Erythroid Cells, *The Journal of Biological Chemistry*, 279(7):5480–5487
- Tang, Y., Liu, D.P., Liang, C.C. (2002) Further understanding of the β -globin locus regulation at the molecular level: looping or linking models?, *Genes to Cells*, 7:889-900
- Thein, S.L. (2004) Genetic insights into the clinical diversity of β -thalassaemia, *British Journal of Haematology*, 124:264–274
- Zhao, Q., Cumming, H., Cerruti, L., Cunningham, J.M., Jane, S.M. (2004) Site-specific Acetylation of the Fetal Globin Activator NF-E4 Prevents Its Ubiquitination and Regulates Its Interaction with the Histone Deacetylase, HDAC1, *The Journal of Biological Chemistry*, 279(40):41477–41486

