Role of Shwachman-Bodian-Diamond syndrome protein in translation machinery and cell chemotaxis: a comparative genomics approach

Olga Vasieva
Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom; Fellowship for the Interpretation of Genomes, Burr Ridge, IL, USA

Abstract: Shwachman-Bodian-Diamond syndrome (SBDS) is linked to a mutation in a single gene. The SBDS protein involved in RNA metabolism and ribosome-associated functions, but SBDS mutation is primarily linked to a defect in polymorphonuclear leukocytes unable to orient correctly in a spatial gradient of chemoattractants. Results of data mining and comparative genomic approaches undertaken in this study suggest that SBDS protein is also linked to tRNA metabolism and translation initiation. Analysis of crosstalk between translation machinery and cytoskeletal dynamics provides new insights into the cellular chemotactic defects caused by SBDS protein malfunction. The proposed functional interactions provide a new approach to exploit potential targets in the treatment and monitoring of this disease.

Keywords: Shwachman-Bodian-Diamond syndrome, wybutosine, tRNA, chemotaxis, translation, genomics, gene proximity

Introduction
Shwachman-Bodian-Diamond syndrome (SBDS) is an autosomal recessive disorder characterized by pancreatic exocrine insufficiency, bone marrow dysfunction, and skeletal abnormalities.1

The SBDS gene (sbds) encodes a member of a highly conserved protein family of unknown function, with orthologs in diverse species, including Archaea, plants, and eukaryotes, but not Eubacteria.2 Structural and functional aspects of SBDS link this protein to a group of proteins involved in RNA metabolism with ribosome-associated functions.3,4 For example, the yeast SBDS ortholog clusters with RNA-processing enzymes and ribosomal RNA-processing factors in gene expression analyses,5,6 and in the SBDS sequence homolog in Saccharomyces cerevisiae, YLR022C, was shown to be physically associated with proteins involved in ribosome biosynthesis.7

The N-terminal domain of the SBDS protein, which is most prone to mutations leading to disease, contains a novel mixed α/β fold that was also suggested for a single domain yeast protein, Yhr087wp, implicated in RNA metabolism.1 The βαββαβ folding topology of the SBDS C-terminal domain III (Glu162–Gly234) is structurally close to domain V of S. cerevisiae elongation factor 2 (Eft2p/Ydr385wp).8 It is also typical for a ferredoxin-like fold and characteristic of a number of heterogeneous nuclear ribonucleoproteins, eg, factors involved in regulation of alternative splicing.5 Although SBDS links to RNA metabolism are apparent, an association has also been observed between the SBDS protein and cell locomotion, in that polymorphonuclear leukocytes isolated from SBDS patients are unable to orient correctly in a
Spatial gradient of chemoattractants. SBDS protein was also shown to localize to the pseudopod of Dictyostelium amoebae during chemotaxis. The interaction of SBDS and a structural or regulatory cytoskeletal component is more than likely responsible for the observed defect in polymorphonuclear leukocyte chemotaxis. Nonetheless, no specific candidate for such an interaction has been suggested.

The complexities of the myriad defects associated with SBDS have made it difficult to relate the diverse biochemical and phenotypic properties of the SBDS syndrome on an experimental basis. A way forward on how mutations of the SBDS gene (sbds) may relate to its altered functions is via a functional genomics approach and phylogenetic analysis of gene proximity in sbds loci. Functionally related genes are commonly found clustered in prokaryotic and eukaryotic genomes, and predicting gene function based on physical proximity to other genes has been used successfully in a number of studies. Therefore, we treated consistency of gene proximity in sbds loci in evolutionary distant genomes as an indication of functional relatedness, which led to a prediction of SBDS protein involvement in initiation of translational wybutosine metabolism. The crosstalk between the translation machinery and elements of the cytoskeleton provides an explanation as to how cell chemotactic defects may be caused by SBDS malfunction.

Materials and methods
We used the Seed database (http://theseed.uchicago.edu/FIG/index.cgi) for chromosome alignment and phylogenetic analysis of gene positional clusters. The “Compare Regions” resource provided by Seed allows alignment of chromosome loci that contain open reading frames for homologous proteins, or, in other words, to pin these loci through genes that are homologous to a query sequence. It can be used in a text or graphic format. We used the latter to illustrate phylogenetic conservation of gene proximity. The typical graphic window presents a selected number of chromosome loci from different genomes. The first line of Compare Regions is a graphical display of the chromosomal neighborhood of the features in its genome. All proteins are shown as colored arrows, where the direction depicts the strand of the feature. RNAs and other features are small boxes on the line. Feature overlaps are resolved by drawing the overlapping feature in a new line. The graph is centered on the selected feature, always numbered 1 and colored red. Below, there is the same region for orthologs in other organisms, also colored in red. The colors of the other features (as well as the numbers) also represent ortholog (or sometimes also paralog) features.

Whenever there are at least two ortholog or paralog features of a kind, a color (and a number) is assigned to them. The selection of genomes to show in the graphics can be made by similarity or the pair of close homologs pin. We used similarity, which means that the genomes are chosen using the similarity of the selected genes to its orthologs in other genomes. The E value cutoff for selection of pinned coding sequence depicts the minimum similarity in order for its region to be displayed. We used the e-20 E value threshold to obtain all the presented data sets.

There are numerous queering and display options that allow customization of the size of displayed regions, selection of organisms, similarity thresholds for pinning of regions, and coloring of features that we implemented to deliver the illustrations accompanying this paper.

Results
Conservation of gene proximity in SBDS gene loci
Phylogenetic analysis of archeal sbds loci (Figure 1) shows conservation of sbds gene proximity. Sbds orthologs are shown as red arrows (N1) in the centers of all the selected regions, where one can also see repetitive occurrence of colors/numbers depicting other orthologous genes in different genomes. Almost all of these co-occurring genes are related to RNA modification and degradation, ie, probable exosome complex exonuclease 2 (EC 3.1.13.-)/tRNA nucleotidyltransferase (N2), proteasome subunit α (EC 3.4.25.1) (N3), probable exosome complex RNA-binding protein 1 (N4), large ribosomal subunit protein L37 Ae (N5) large ribosomal subunit protein L15e (N7), ribonuclease P (tRNA processing) protein component 3 (EC 3.1.26.5) (N8), ribonuclease P protein component 2 (EC 3.1.26.5) (N9), prefoldin, chaperonin cofactor (N10), and a predicted exosome subunit containing the IMP4 domain present in small nuclear ribonucleoprotein (N11). An archeal sbds locus that includes all or part of the genes encoding the listed functions is surrounded by a variable region (gray arrows), suggesting that clustered genes related to the archeal exosome complex indeed represent a functionally coupled group or even an operon, and that sbds can be a part of this group.

Chromosome loci with vertebrate orthologs to human sbds (ENSG00000126524) have also been aligned. In the human genome, sbds spans a start for tyw1 (ENSG00000198874), which encodes tRNA-wybutosine synthesizing protein (TYW1) from the opposite string. In all vertebrate genomes analyzed (from the SEED database collection), sbds (red, N1)
is colocalized with a gene encoding the TYW1 homolog, annotated in SEED as wybutosine biosynthesis reductase (green, N2, Figure 2). A small gap between these two genes in the different vertebrate genomes suggests that they may share a regulatory region. The gene region downstream of \textit{sbds} is not conserved, even in narrower phylogenetic groups, because one can see it for the displayed regions from several mammalian genomes.

In the \textit{Plasmodium falciparum} and \textit{Ciona intestinalis} genomes, \textit{sbds} genes are also colocalized with genes encoding functions related to translational initiation and mainly with tRNA-regulated components of translational initiation complexes, i.e., a gene for predicted GCN1 (general control of amino-acid synthesis 1) (XP\_002120729.1) is immediately downstream and a gene for tRNA/RNA cytosine-C5-methylase (EC 2.1.1.-) is immediately upstream from the human \textit{sbds} ortholog (XM\_002124782.1) in the \textit{C. intestinalis} genome (Table 1). The \textit{aIF2β} gene is the second gene upstream from the \textit{sbds} ortholog (XP\_001348280.1) in a \textit{Plasmodium} genome (Table 2).

In \textit{Schizosaccharomyces pombe}, the \textit{sbds} ortholog (O14179, SDO1) is proximal to a gene encoding the L26e ribosomal protein, followed by a gene for the mRNA capping enzyme subunit. Gene encoding RNA processing factor 1 is immediately downstream, followed by one for S23e (Table 3).

No specific functional gene overrepresentation at the \textit{sbds} locus in \textit{S. cerevisiae} was observed, apart from a noticeable but comparatively distant proximity to genes...
encoding RNA-related functions. However, sbds orthologs (XP_387070.1, XP_956795.1, XP_360288.1) are proximal to genes encoding L18e in fungal (Gibberella zeae, Neurospora crassa, and Magnaporthe oryzae) genomes (Figure 3).

Discussion

SBDS is a complex disease linked to mutation in a single gene. Recent advances in genomics and functional bioinformatics are providing a new avenue for studying this and similar phenomena, complementing traditional hypothesis-driven laboratory research. Our bioinformatics approach undertaken in this study provides a new insight as to how a single gene mutation may affect diverse molecular functions, such as RNA metabolism, translation initiation, and cytoskeletal dynamics leading to complex disease.

Phylogenetic conservation of gene proximity may manifest itself in the presence of sets of orthologous genes in gene loci from different genomes. However, even in prokaryotes, it often comes up as conservation of positional connections between and inside particular pathways, not exactly the same representatives of the pathways. Therefore, integration of genomic data may uncover trends in gene positional clustering and the underlying functional links between genes and pathways. Analysis of our data points to initiation of translation as the main functional
connection for SBDS (Figure 3), with particular TYW1 and SBDS functional coupling as characteristic of all the studied vertebrate genomes.

Wybutosine is a hypermodified guanosine with a tricyclic base found at the 3′-position adjacent to the anticodon of eukaryotic phenylalanine tRNA. The UUU phenylalanine codon is highly prone to frameshift in the 3′ direction at pyrimidine 3′ contexts, and wybutosine supports reading frame maintenance by stabilizing codon–anticodon interactions during decoding on the ribosome.

Wybutosine synthesis might proceed through sequential reactions in a multiple protein complex assembled with the precursor tRNA and may be linked to exosomal/proteosomal structure that includes the SBDS protein. Comparative analysis of archael sbsd loci stresses the potential involvement of SBDS in the exosomal complex where functions of translation, RNA processing, and degradation are tightly coupled.

The SBDS protein has a thiorexin fold, which might be directly coupled to the function of wybutosine reductase. Not all stages of wybutosine biosynthesis are known. There is some ambiguity surrounding the source of the C2 atom in wybutosine, suggested to originate from an intermediate in lysine metabolism. Interestingly, SBDS was shown to interact physically (Reactome db) with methylmalonate semialdehyde dehydrogenase, the enzyme in the branched-chain amino acid degradation pathway. It would be too speculative to go further, but the possible involvement of the SBDS protein in biosynthesis of wybutosine merits experimental validation.

Considering an alternative link between wybutosine and SBDS, we may suggest involvement of SBDS in degradation of noncharged or not properly modified tRNA. The tRNA surveillance pathway has been shown to exist in yeast and requires the exosome for polyadenylation and degradation of hypomodified pre-tRNA(i)(Met).

Analogously, tRNA(Phe) without a modified wybutosine residue may be subject to degradation, and therefore SBDS may play some role in this process. The N-terminus of the Yhr087wp yeast protein has a fold similar to that in SBDS. It also has 2.4e-06 and 100.00% homology with bacterial RNA pseudouridine 13 synthase, that may support an involvement of the SBDS protein in tRNA modification. Genes do not cluster in S. cerevisiae orthologs of SBDS (Q07953, SDO1) and TYW1 (Q08960, Tyw1p). However, the tyw1 locus in S. cerevisiae also contains a gene for tRNA pseudouridine synthase 1 (YPL212C) and a number of genes encoding proteins of ribosomal biogenesis.

Establishing a link between SBDS and tRNA modification provides an interesting aspect of crosstalk between cytoskeleton function and translation machinery that may explain a faulty chemotactic phenotype of SBDS malfunction.

### Table 2: Tabular display of the chromosomal neighborhood of sbsd gene in Plasmodium falciparum genome (SEED db IDs)

<table>
<thead>
<tr>
<th>Fig</th>
<th>ID</th>
<th>ID</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>36329.1.peg.3840</td>
<td>2666151</td>
<td>2666990</td>
<td>+</td>
</tr>
<tr>
<td>36329.1.peg.3841</td>
<td>2667934</td>
<td>2667270</td>
<td>–</td>
</tr>
<tr>
<td>36329.1.peg.3842</td>
<td>2669311</td>
<td>2670276</td>
<td>+</td>
</tr>
<tr>
<td>36329.1.peg.3843</td>
<td>2671220</td>
<td>2670798</td>
<td>–</td>
</tr>
<tr>
<td>36329.1.peg.3844</td>
<td>2673717</td>
<td>2671936</td>
<td>–</td>
</tr>
</tbody>
</table>

**Note:** Line corresponded to a gene for SBDS protein ortholog is highlighted.

### Table 3: Tabular display of the chromosomal neighborhood of SBDS-encoding genes in Schizosaccharomyces pombe genome (SEED db IDs)

<table>
<thead>
<tr>
<th>Fig</th>
<th>ID</th>
<th>ID</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>4896.1.peg.3836</td>
<td>2660945</td>
<td>2660505</td>
<td>–</td>
</tr>
<tr>
<td>4896.1.peg.3837</td>
<td>2661984</td>
<td>2662340</td>
<td>+</td>
</tr>
<tr>
<td>4896.1.peg.3838</td>
<td>2666320</td>
<td>2662700</td>
<td>–</td>
</tr>
<tr>
<td>4896.1.peg.3839</td>
<td>2665276</td>
<td>2664342</td>
<td>–</td>
</tr>
<tr>
<td>4896.1.peg.3840</td>
<td>2666151</td>
<td>2666990</td>
<td>+</td>
</tr>
<tr>
<td>4896.1.peg.3841</td>
<td>2667934</td>
<td>2667270</td>
<td>–</td>
</tr>
<tr>
<td>4896.1.peg.3842</td>
<td>2669311</td>
<td>2670276</td>
<td>+</td>
</tr>
<tr>
<td>4896.1.peg.3843</td>
<td>2671220</td>
<td>2670798</td>
<td>–</td>
</tr>
<tr>
<td>4896.1.peg.3844</td>
<td>2673717</td>
<td>2671936</td>
<td>–</td>
</tr>
</tbody>
</table>

**Note:** Line corresponded to a gene for SBDS protein ortholog is highlighted.
In the *P. falciparum* and *C. intestinalis* genomes, SBDS genes are colocalized with functions associated with tRNA-regulated components of translational initiation complexes.\(^{24}\) GCN1 is a translational activator and regulator of Gcn2p kinase activity. It forms a complex with Gcn20p and is proposed to stimulate Gcn2p activation by an uncharged tRNA.\(^{24,25}\) The second gene colocalized with *sbds* in the sea squin genome is a gene for tRNA/RNA cytosine C5 methylase, which is required for initiating tRNA(i) (Met) synthesis. Eukaryotic and archaean initiation factors 2 are heterotrimeric proteins where only the \(\gamma\) subunit of the \(\alpha\)\(\beta\) heterodimer contacts tRNA.\(^{26}\) Intriguingly, the C5 methylase, which is required for initiating tRNA(i) (Met) synthesis, is separated from cell migration caused in primitive eukaryotic ancestors by starvation or by an avoidance response to other stressors.

There is also a crosstalk between the two systems in establishing cell polarity during chemotaxis. It has been proposed that the EF-1\(\alpha\)-F-actin complex is an important scaffold for anchoring of \(\beta\)-actin mRNA to sites of active actin polymerization.\(^{27}\) Translation only occurs when the RNA-protein complex reaches its destination at the cell periphery where actin translation and assembly takes place.\(^{27}\) F-actin and aa-tRNA compete for EF-1\(\alpha\), and their binding is \(\mathrm{pH}\)-dependent and mutually exclusive. Release of EF-1\(\alpha\) from actin binding was suggested to cause a transient increase in local concentration of the factor to facilitate polypeptide elongation. This interrelationship may ensure that cell proliferation and steady-state protein synthesis is separated from cell migration caused in primitive eukaryotic ancestors by starvation or by an avoidance response to other stressors.

Indirect evidence implicates actin as a cofactor in eukaryotic protein synthesis. The principle function of EF-1\(\alpha\) is to bind aminoacyl-tRNA to the ribosome. EF-1\(\alpha\) also interacts with the cytoskeleton by binding and bundling actin filaments and microtubules, and can alter the assembly of F-actin, a filamentous scaffold on which nonmembrane-associated protein translation take place\(^{27}\) (Figure 4). F-actin and aa-tRNA compete for EF-1\(\alpha\), and their binding is \(\mathrm{pH}\)-dependent and mutually exclusive. Release of EF-1\(\alpha\) from actin binding was suggested to cause a transient increase in local concentration of the factor to facilitate polypeptide elongation. This interrelationship may ensure that cell proliferation and steady-state protein synthesis is separated from cell migration caused in primitive eukaryotic ancestors by starvation or by an avoidance response to other stressors.

Figure 3  Graphical display of the chromosomal neighborhood of *sbds* genes in fungal genomes. Arrows correspond to open reading frames encoding: 2-SSU ribosomal protein S18e (S13p), 3-carboxypeptidase precursor, and 4- and 5-hypothetical proteins. 100 kbp regions are shown. Organisms are *Neurospora crassa*, *Gibberella zeae*, *Plasmodium falciparum*, and *Magnaporthe grisea*.

![Image](https://www.dovepress.com/www.dovepress.com/submit-your-manuscript.png)

**Figure 4**  Hypothetical functions of SBDS protein, and their effect on chemotaxis. Clouds represent gene proximity in different groups of organisms, ie, vertebrates (yellow), *Ciona intestinalis* (blue), and *Plasmodium falciparum* (purple).
and cell locomotion. Experimental validation of the relationship between SBDS, wybutosine synthesis, and/or degradation of tRNA_{phe} seems plausible, and we hope that our suggestions will attract the attention of biologists in related fields.

It is still not clear what functional features of a gene pair (structural or functional specificity of the encoded proteins, topology of their interaction, presence of a direct protein–protein contact) correlate significantly with their colocalization. Validation of genomic clustering of genes encoding metabolic functions demonstrates 90% correlation for yeast and just slightly less for a human genome. A high correlation was also shown between gene colocalization and their temporal and spatial expression profiles. 14,16 All the existing studies point to phenotypic associations between genes clustered in genomes, but more information is required for a proper large-scale statistical analysis of the correlation. We hope that case-by-case analysis will support the general validity of this method and lead to a routine automatic approach to functional classification of eukaryotic proteins via systematic comparative genomics.

Acknowledgments
The author thanks Dr Bahram Ibrahimi for his useful comments and also colleagues at the Fellowship of Interpretation of Genomes for their constant support and inspiration.

Disclosure
The author reports no conflict of interest in this work.

References

