



RESEARCH ARTICLE

A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling. [version 1; referees: 2 approved]

Reinier Gesto-Borroto, Miriam Sánchez-Sánchez, Raúl Arredondo-Peter

Laboratorio de Biofísica y Biología Molecular, Centro de Investigación en Dinámica Celular, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Colonia Chamilpa, Morelos, 62210, Mexico

v1 First published: 13 May 2015, 4:117 (doi: [10.12688/f1000research.6392.1](https://doi.org/10.12688/f1000research.6392.1))
 Latest published: 13 May 2015, 4:117 (doi: [10.12688/f1000research.6392.1](https://doi.org/10.12688/f1000research.6392.1))

Abstract

Globins (Glbs) are proteins widely distributed in organisms. Three evolutionary families have been identified in Glbs: the M, S and T Glb families. The M Glbs include flavohemoglobins (fHbs) and single-domain Glbs (SDgbs); the S Glbs include globin-coupled sensors (GCSs), protoglobins and sensor single domain globins, and the T Glbs include truncated Glbs (tHbs). Structurally, the M and S Glbs exhibit 3/3-folding whereas the T Glbs exhibit 2/2-folding. Glbs are widespread in bacteria, including several rhizobial genomes. However, only few rhizobial Glbs have been characterized. Hence, we characterized Glbs from 62 rhizobial genomes using bioinformatics methods such as data mining in databases, sequence alignment, phenogram construction and protein modeling. Also, we analyzed soluble extracts from *Bradyrhizobium japonicum* USDA38 and USDA58 by (reduced + carbon monoxide (CO) minus reduced) differential spectroscopy. Database searching showed that only *fhb*, *sdgb*, *gcs* and *thb* genes exist in the rhizobia analyzed in this work. Promoter analysis revealed that apparently several rhizobial *glb* genes are not regulated by a -10 promoter but might be regulated by -35 and Fnr (fumarate-nitrate reduction regulator)-like promoters. Mapping analysis revealed that rhizobial *fhbs* and *thb* s are flanked by a variety of genes whereas several rhizobial *sdgbs* and *gcss* are flanked by genes coding for proteins involved in the metabolism of nitrates and nitrites and chemotaxis, respectively. Phenetic analysis showed that rhizobial Glbs segregate into the M, S and T Glb families, while structural analysis showed that predicted rhizobial SDgbs and fHbs and GCSs globin domain and tHbs fold into the 3/3- and 2/2-folding, respectively. Spectra from *B. japonicum* USDA38 and USDA58 soluble extracts exhibited peaks and troughs characteristic of bacterial and vertebrate Glbs thus indicating that putative Glbs are synthesized in *B. japonicum* USDA38 and USDA58.



This article is included in the [Oxygen-binding and sensing proteins](#) channel.

Open Peer Review

Referee Status:

Invited Referees

1 2

version 1

published
13 May 2015



report



report

1 **Manuel Becana**, Spanish National Research Council Spain

2 **Paul Twigg**, University of Nebraska at Kearney USA

Discuss this article

Comments (0)

Corresponding author: Raúl Arredondo-Peter (ra@uaem.mx)

How to cite this article: Gesto-Borroto R, Sánchez-Sánchez M and Arredondo-Peter R. **A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling.** [version 1; referees: 2 approved] *F1000Research* 2015, 4:117 (doi: [10.12688/f1000research.6392.1](https://doi.org/10.12688/f1000research.6392.1))

Copyright: © 2015 Gesto-Borroto R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was partially financed by SEP-PROMEP (grant number UAEMor-PTC-01-01/PTC23) and Consejo Nacional de Ciencia y Tecnología (CoNaCyT grant numbers 25229N and 42873Q), México. R. Gesto-Borroto is a graduate student financially supported by CoNaCyT (registration no. 293307).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 13 May 2015, 4:117 (doi: [10.12688/f1000research.6392.1](https://doi.org/10.12688/f1000research.6392.1))

Introduction

Globins (Glbs) are proteins widely distributed in organisms from the three kingdoms of life, *i.e.* in Archaea, Eubacteria and Eukarya¹. Structurally, Glbs fold into a tertiary structure known as the globin fold. This protein folding consists of six to eight α -helices (designated with letters A to H) that form a hydrophobic pocket where a heme prosthetic group is located². Two structural types of the globin fold have been identified in Glbs: the 2/2- and 3/3-fold. In the 2/2-Glbs, helices B and E overlap to helices G and H³ and in the 3/3-Glbs helices A, E and F overlap to helices B, G and H^{4,5}. Likewise, three evolutionary families have been identified in Glbs^{6,7}: the M, S and T Glb families. The M Glbs include flavohemoglobins (fHbs) and single-domain Glbs (SDgbs), the S Glbs include globin-coupled sensors (GCSs), protoglobins and sensor single domain globins, and the T Glbs include truncated Glbs (tHbs) (which are further classified into class 1, class 2 and class 3 tHbs). Canonical tHbs are ~20 to 40 amino acids shorter than the globin fold, resulting in an almost absent helix A and a helix F that is reduced to a single turn^{8,9}. The M and S Glbs fold into the 3/3-fold whereas the T Glbs fold into the 2/2-fold.

A variety of gaseous ligands bind to the heme Fe of Glbs, most notably O₂ and nitric oxide (NO). The reversible binding of O₂ is associated with the major function of Glbs in organisms: the transport of O₂. Binding of NO by oxygenated Glbs is essential to NO-detoxification *via* NO-dioxygenase activity^{10,11}. Several additional functions have been reported for Glbs, including dehaloperoxidase activity and reaction with free radicals, binding and transport of sulfide and lipids, and O₂-sensing (reviewed by Giardina *et al.*¹² and Vinogradov *et al.*¹³). This indicates that *in vivo*, Glbs might be multifunctional proteins.

Glbs are widespread in bacteria. A comprehensive genomic analysis revealed that *glb* genes belonging to the M, S and T Glb families exist in the genomes of 1185 Eubacteria, including several rhizobial genomes⁷. However, only few rhizobial *glb* genes have been characterized. Characterizing rhizobial Glbs is of interest because rhizobia establish symbiotic relationships with leguminous plants. A result of this plant-microbe interaction is the symbiotic fixation of atmospheric N₂, which occurs within specialized plant organs called nodules¹⁴. Symbiotic N₂-fixation is a process modulated by a variety of factors, such as the O₂¹⁵ and NO^{16,17} levels in the surrounding environment. Glbs bind O₂ and NO and thus may function in some aspects of the N₂-fixation, *e.g.* by transporting O₂ and detoxifying NO. Modulation of O₂ levels in the plant cell cytoplasm from nodules is well characterized^{18,19}. A plant Glb (leghemoglobin (Lb)) that is synthesized at high (~3 to 5 mM) concentrations in nodules apparently facilitates O₂-diffusion to the symbiotic rhizobia and maintains low (submicromolar) concentrations of O₂ within nodules. This is essential for sustaining the (micro) aerobic respiration of symbiotic rhizobia and preventing the inactivation of nitrogenase (which fixes the atmospheric N₂ into NH₄⁺) by O₂. The binding and metabolizing of NO by Lb and other Glbs is also well documented^{11,20}. Thus, a likely function for Lb in nodules is to detoxify the NO that is generated during the plant infection by rhizobia²¹. However, little is known about the properties and functions of Glbs either within the symbiotic or free-living rhizobia.

Forty-six years ago Appleby²² was the first to propose the existence of Glbs in rhizobia. This author detected absorption peaks and troughs that are characteristic of Glbs in differential (dithionite reduced + CO minus dithionite reduced) spectra of soluble extracts from *Bradyrhizobium japonicum* 505 (Wisconsin). Subsequent spectroscopic analyses suggested the existence of soluble Glbs in *Rhizobium leguminosarum* *bv. viciae*²³, *B. japonicum* NPK63²⁴ and *R. etli* CE3²⁵. The first rhizobial *glb* gene was identified in the pSymA megaplasmid of *Sinorhizobium meliloti* 1021²⁶. BLAST analysis revealed that this gene corresponded to an *fhb* gene and thus was named *smfhb*. A bioinformatics analysis showed that *smfhb* is flanked by *nos* and *fix* genes (which code for denitrification enzymes and high O₂-affinity terminal oxidases and an O₂-sensor, respectively) and that apparently it is regulated by an Fnr-like promoter. These observations suggested that *smfhb* is regulated by the concentration of O₂ and that SmfHb functions in some aspects of nitrogen metabolism. A transcriptomic analysis of the *S. meliloti* response to NO in culture showed that *smfhb* (also designated as a *S. meliloti hmp*) is upregulated by NO and the analysis of a *smfhb* mutant exhibited a high sensitivity to NO in culture and led to a reduced N₂-fixation efficiency *in planta*. These observations suggested that SmfHb functions in some aspects of NO metabolism, possibly by detoxifying NO²⁷.

Genomic analysis reported by Vinogradov *et al.*⁷ revealed that Glb sequences exist in several rhizobia. However, in spite of the above reports knowledge on the rhizobial Glbs is quite limited. Hence, in order to obtain information on the properties of rhizobial Glbs we characterized Glb sequences from selected rhizobial genomes by using bioinformatics methods. These included gene characterization, polypeptide sequence and phenetic analysis, as well as protein modeling. Also, we analyzed soluble extracts from *B. japonicum* USDA38 and USDA58 by differential spectroscopy. Our main results showed that only *fhb*, *sdgb*, *gcs* and *thb* genes exist in the rhizobia analyzed in this work; that several rhizobial *glb* genes are not regulated by a -10 promoter but might be regulated by -35 and Fnr-like promoters; that rhizobial *fHbs* and *tHbs* are flanked by a variety of genes whereas several rhizobial *sdgbs* and *gcss* are flanked by genes coding for proteins involved in the metabolism of nitrates and nitrites and chemotaxis, respectively; that rhizobial Glbs segregate into the M, S and T Glb families; that predicted rhizobial SDgbs and fHbs and GCSs globin domain and tHbs fold into the 3/3- and 2/2-fold, respectively, and that spectra from *B. japonicum* USDA38 and USDA58 soluble extracts exhibit peaks and troughs characteristic of bacterial and vertebrate Glbs.

Methods

Database search

Putative Glb sequences and Glb domains were identified in databases (Table S1) containing the genomes of rhizobial species and strains using the query sequences *S. meliloti* fHb; *Vitreoscilla* SDgb; *Agrobacterium tumefaciens* GCS; *Methanosarcina acetivorans* protoglobin; *Methylacidiphilum inferorum* sensor single domain globin; *Mycobacterium tuberculosis* tHb class 1; *A. tumefaciens* tHb class 2, and *M. avium* tHb class 3 (Genbank accession numbers AY328026, AAA75506, NP_354049, 2VEB_A, YP_001939425, NP_216058, WP_020813663 and BAN32501, respectively) and the

SUPERFAMILY database (<http://supfam.mrc-lmb.cam.ac.uk>)²⁸. Resulting sequences were subjected to a FUGUE analysis (<http://tardis.nibio.go.jp/fugue/prfsearch.html>)²⁹ to determine the most similar Glb structure and presence of proximal H at the myoglobin-fold position F8. Putative Glbs had to satisfy the following criteria: length higher than or ~100 amino acids, a FUGUE Z score higher than 6 (which corresponds to 99% specificity²⁹) with known Glb structures, and the presence of proximal H at position F8.

Gene mapping and detection of promoter sequences

Scaffolds containing copies of the *glb* gene were used for mapping *glbs*. This included the detection of open reading frames (ORFs) ~5 kb up- and downstream to *glbs* and ORF length, transcription direction and localization in the +/- strand. Canonical (-10 and -35) and Fnr³⁰ promoter sequences and Shine-Dalgarno sequences were searched within 130 nucleotides upstream to the rhizobial *glb* genes either by using the search tool of MS Word® or by pairwise sequence alignments using the ClustalX program (<http://www.clustal.org/clustal2/>)³¹.

Protein sequence alignments and phenetic analysis

Pairwise and multiple sequence alignments were performed using the ClustalX program³¹. Multiple sequence alignment was manually verified using the procedure described by Kapp *et al.*³² based on the myoglobin-fold³³. A phenogram was constructed from the aligned sequences using the UPGMA method from the ClustalX program. The resulting phenogram was edited using the iTOL program (<http://itol.embl.de/>)³⁴.

Modeling and analysis of the predicted proteins tertiary structure

The tertiary structure of rhizobial Glbs was modeled using the automated mode of the I-TASSER server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>)³⁵⁻³⁷, which also provided the best structural homologs to the query sequences. Models were edited using the VMD program (<http://www.ks.uiuc.edu/Research/vmd/>)³⁸ and Adobe Photoshop® software. Distance and dihedral angles of amino acids at the heme prosthetic group were calculated using the distance and dihedral tools of the SwissPDBViewer program (<http://spdbv.vital-it.ch/>) as described by Gopalasubramaniam *et al.*³⁹ and Sáenz-Rivera *et al.*⁴⁰, respectively.

Bacterial growth, cell rupture and spectral analysis

Bradyrhizobium japonicum USDA38 and USDA58 were kindly provided by Drs. Donald Keister and Douglas Jones (United States Department of Agriculture, USA). All reagents were purchased from Sigma-Aldrich (St. Louis MO, USA). *B. japonicum* cells were grown in YM (Yeast Mannitol) broth (per 100 ml: KH₂PO₄, 50 mg; MgSO₄, 20 mg; NaCl, 10 mg; mannitol, 1 g; yeast extract, 50 mg, pH 7.0) for 3 to 5 days at 30°C with shaking at 200 rpm. Cells were harvested by centrifugation at 11,000 × g, pellets were resuspended in 50 mM Na-phosphate buffer (pH 7.2) containing 1 mM EDTA and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were disrupted by sonication at maximum power (three cycles of 1 min each in ice) and incubation at 4°C overnight with gentle agitation after the addition of DNase I (40 U/ml), RNase A (3 U/ml) and lysozyme (2 mg/ml). The resulting solution was cleared by centrifugation at 22,000 × g for 40 min at 4°C, and the supernatant was fractionated

with solid ammonium sulphate between 35 and 65% saturation. The resulting pellet was resuspended in 5 ml of 50 mM Na-phosphate buffer (pH 7.2) containing 1 mM EDTA and 1 mM PMSF and dialyzed for 18 h against the same buffer to remove the excess of salts. 0.5 to 1 ml aliquots of the dialyzed solution were used to obtain the dithionite reduced + CO minus dithionite reduced differential spectra in a Beckman DU6 spectrophotometer. Control spectra were obtained from commercial (Sigma-Aldrich) preparations of the sperm whale myoglobin and bovine blood hemoglobin.

Dataset 1. Globin genes detected in the genomes of rhizobial bacteria

<http://dx.doi.org/10.5256/f1000research.6392.d46189>

Globin nomenclature corresponds to the first three binomial (genus and species) letters followed by the strain name, globin type and gene copy number. URLs indicate links to individual *glb* gene sequences⁵⁶.

Dataset 2. Predicted Glb polypeptides detected in the genomes of rhizobial bacteria

<http://dx.doi.org/10.5256/f1000research.6392.d46190>

Globin nomenclature corresponds to the first three binomial (genus and species) letters followed by the strain name, globin type and globin copy number. URLs indicate links to individual Glb polypeptide sequences⁵⁷.

Dataset 3. Distance to the heme Fe and orientation of distal, proximal, B10 and CD1 amino acids in the predicted structure of selected rhizobial Glbs (Table S2)

<http://dx.doi.org/10.5256/f1000research.6392.d46191>

Structural homologs (including the PDB ID number), amino acids from the structural homologs and values for the structural homologs amino acids to individual rhizobial Glbs are indicated in parenthesis for comparison⁵⁸.

Results and discussion

Detection of Glb sequences in the genomes of α - and β -rhizobia

Recently, Vinogradov *et al.*⁷ reported that Glb sequences exist in the genomes of 96 rhizobia. However, this report did not provide the rhizobial Glb sequences or links to rhizobial scaffolds containing the Glb sequences. Hence, we searched in databases (see the Methods section and Table S1) in order to obtain rhizobial Glb sequences for analysis. We selected 62 out of the 96 rhizobial genomes reported by the above authors representing the major rhizobial genera, species and strains, which included α - and β -rhizobia (*i.e.* those classified within the α - and β -proteobacteria, respectively). A total of 197 *glb* sequences were detected in the 62 rhizobial genomes, corresponding to 7 *fhbs*, 47 *sdgbs*, 40 *gcss* and 103 *thbs* (4 *thbs* class 1, 56 *thbs* class 2 and 43 *thbs* class 3). Individual Glb nucleotide and polypeptide sequences and links to rhizobial scaffolds containing the Glb sequences are provided in Dataset 1 and Dataset 2, respectively. All the rhizobial genomes analyzed in this work contained

glb sequences, thus indicating that *glbs* are widespread in rhizobia. However, protoglobin and sensor single domain globin sequences were not detected in the rhizobial genomes. This observation indicates that apparently only the *fhb*, *sdgb*, *gcs* and *thb* lineages evolved within rhizobia.

A distribution analysis showed that most (61) of the rhizobial genomes analyzed in this work contain *thbs*, either as single *thbs* (13) or in combination with *fhbs*, *sdgbs* and/or *gcss* (48). Furthermore, one rhizobial genome contained only a *gcs* and none contained only *fhbs* and *sdgbs* and the combinations *fhbs* + *sdgbs*, *fhbs* + *gcss* and *sdgbs* + *gcss* (Figure 1). These observations indicate that in the rhizobia analyzed in this work *thbs* predominate over other *glbs* and that in these bacteria *fhbs*, *sdgbs* and *gcss* mostly exist in combination with *thbs*. Also, analysis of the *glb* copy number showed that in the rhizobia analyzed in this work *fhbs* mostly exist as single copy (ranging from one to two copies), *sdgbs* mostly exist as two copies (ranging from one to four copies), *gcss* exist as either single or two copies (ranging from one to two copies) and *thbs* mostly exist as two copies (ranging from one to three copies) although quite a few *thbs* exist as single copy (Table 1). Thus, apparently rhizobial *glbs* mostly exist as either single or two copies.

Mapping of *glb* genes in the rhizobial genomes

The *glb* genes detected in this work were mapped within the rhizobial genomes in order to identify genes that flank nearby to and could coexpress with *glbs*. Mapping analysis showed that rhizobial *glb* copies are located in different scaffolds and that they are not tandemly arrayed. Figure S1A shows that either no ORFs or ORFs coding for hypothetical or non-identified proteins are located nearby most of the rhizobial *fhb* genes. However, genes coding for the transcriptional regulator NsrR, 2-nitropropane dioxygenase and NosR, Z, D, F, Y and X are located nearby *cupnecN1fhb1*,

rhilegUPM1137fhb and *sinmel1021fhb*, respectively. Figure S1B shows that *B. elkanii* and *B. japonicum* *sdgbs* are mostly flanked by genes coding for proteins that function in nitrate/nitrite metabolism and sugar transport. Figure S1C shows that genes coding for proteins that function in chemotaxis are located nearby several rhizobial *gcss*, although genes coding for a peptide deformylase, sugar and nitrate transport proteins and NAD(P)H nitrate reductase are located nearby some other rhizobial *gcss*. Figure S1D shows

Table 1. Number of *glb* copies detected in the rhizobial genomes analyzed in this work.

<i>glb</i> /no. of copies	No. of genomes
<i>fhbs</i>	
1	5
2	1
<i>sdgbs</i>	
1	5
2	11
3	4
4	2
<i>gcss</i>	
1	12
2	14
<i>thbs</i>	
1	22
2	36
3	3

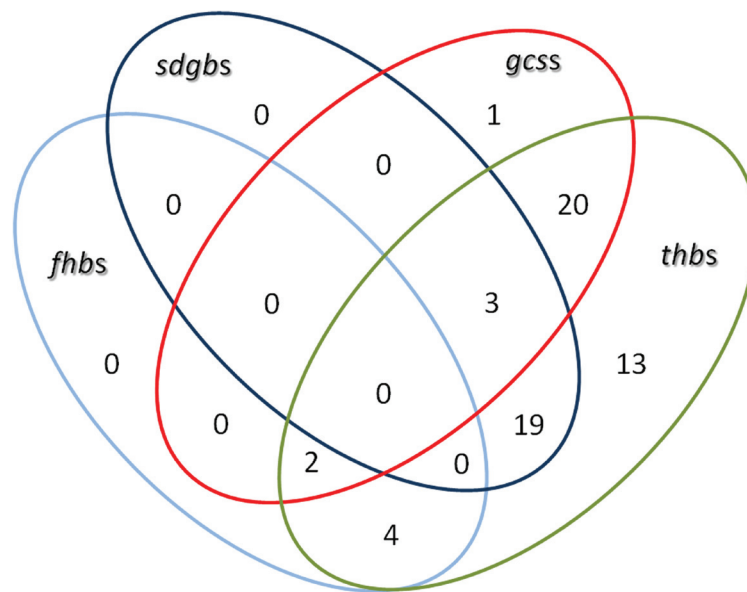


Figure 1. Venn diagram illustrating the distribution of *glb* genes in the rhizobial bacteria analyzed in this work. Numbers correspond to rhizobial genomes containing *glbs*.

that genes flanking the rhizobial *thbs* are rather variable. However, *B. japonicum thbs* are often flanked by genes coding for the transcriptional regulator Rieske Fe-S, shikimate kinase and alcohol dehydrogenase; mesorhizobia *thbs* are often flanked by genes coding for permeases and tRNA-Trp, and *R. leguminosarum thbs* are often flanked by genes coding for membrane proteins. Thus, if *glb* and flanking genes coexpress in rhizobia, and proteins coded by these genes function within the same metabolic pathways, the above observations suggest that rhizobial Glbs could play a variety of roles in rhizobial physiology, including nitrate/nitrite metabolism, transport processes, gene regulation and chemotaxis. Interestingly, with the exception of *sinmel1021fhb* which is flanked by *nos* and *fix* genes (Figure S1A)²⁶, *nif* and *fix* genes coding for proteins that function in N₂-fixation were not detected nearby the rhizobial *glb* genes. This observation suggests that rhizobial Glbs might not directly function in N₂-fixation.

Detection of promoter sequences upstream to the rhizobial *glb* genes

Identification of promoter sequences is crucial to an understanding of gene regulation and ultimately protein function within the

cell's physiology. Hence, we searched for canonical (-10 and -35) promoters and the O₂- and NO-regulated Fnr promoter^{30,41,42} within 130 nucleotides upstream to 44 selected rhizobial *glb* genes (*i.e.* those representative of major rhizobial Glb clades identified in this work (see Figure 2)). Also, we searched for Shine-Dalgarno sequences within the same region, which indicate that Glb transcripts could be translated into proteins. Results showed that, with the exception of *burphySTM815thb1*, *burphySTM815thb2* and *rhilupHPC(L)thb1*, a -10 promoter is absent upstream of the selected rhizobial *glbs*. In contrast, with the exception of *cupnec-N1thb1* and *rhilupHPC(L)thb2*, a -35 promoter exists upstream of the selected rhizobial *glbs*. Searching for Fnr promoter sequences revealed that Fnr-like promoters exist upstream to 30 out of the 44 selected rhizobial *glbs*, including *fhb*, *sdgb*, *gcs* and *thb* genes. A Shine-Dalgarno sequence was detected upstream to most of the selected rhizobial *glbs* (Table 2). These observations suggest that the -35 promoter is a major canonical promoter that regulates most of the rhizobial *glbs*, that it is likely that several rhizobial *glbs* are regulated by levels of O₂ and NO throughout an FNR mechanism⁴¹⁻⁴⁴ and that rhizobial Glb transcripts are translated into proteins.

Table 2. Position of canonical and Fnr-like promoter sequences and Shine-Dalgarno sequence within 130 nucleotides upstream to selected rhizobial *glb* genes. Consensus sequences are indicated in parenthesis. Identical and non-identical nucleotides into the Fnr-like promoter sequences to the consensus Fnr promoter sequence are indicated with upper- and lowercase letters, respectively. N.D., non-detected.

<i>glb</i> gene	Canonical promoters		Fnr promoter		Shine-Dalgarno sequence (AGGAGG)
	-10 promoter (TATAAT)	-35 promoter (TTGACA)	Sequence (TTTAAGAGGCCAAT)	Position	
<i>fhrs</i>					
<i>burphySTM815fhb</i>	N.D.	-36 to -41	TcTAAGcGaCtgAT	-102 to -115	-10 to -13
<i>cupnecHPC(L)fhb</i>	N.D.	-43 to -48	N.D.		-8 to -12
<i>cupnecJMP134fhb</i>	N.D.	-46 to -54	TTTAAaAcGgagcc	-5 to -18	-10 to -15
<i>cupnecN1fhb1</i>	N.D.	-46 to -52	N.D.		-10 to -15
<i>cupnecN1fhb2</i>	N.D.	-29 to -36	aTcAAGgcGgCgAg	-64 to -77	-8 to -12
<i>rhilegUPM1137fhb</i>	N.D.	-32 to -37	N.D.		-9 to -12
<i>sinmel1021fhb</i>	N.D.	-48 to -56	gTcAAGgaGCCAAa	-12 to -25	-8 to -12
			ggTtgGgGtCCAcT	-61 to -74	
<i>sdgbs</i>					
<i>azodoeUFLA1-100sdgb</i>	N.D.	-38 to -42	gccAgGAGtCCgAT	-2 to -15	-8 to -12
<i>braelkUSDA94sdgb2</i>	N.D.	-36 to -43	TaTAAGgacatcAT	-114 to -127	-7 to -11
<i>braelkWSM1741sdgb2</i>	N.D.	-34 to -40	N.D.		-7 to -12
<i>braelkUSDA3254sdgb1</i>	N.D.	-61 to -66	TTTttGgGGCaAAT	-71 to -84	N.D.
<i>braelkUSDA3254sdgb2</i>	N.D.	-40 to -44	TTTAcGAGGctgcT	-11 to -24	-16 to -22
<i>braelkUSDA3259sdgb1</i>	N.D.	-41 to -46	TTTcAGAactCAAT	-22 to -35	-8 to -12
			cTTcgGttaCCAAT	-56 to -69	
<i>brajapUSDA38sdgb2</i>	N.D.	-53 to -58	N.D.		-6 to -9
<i>brajapUSDA124sdgb1</i>	N.D.	-60 to -65	N.D.		-7 to -11

gfb gene	Canonical promoters		Fnr promoter		Shine-Dalgarno sequence (AGGAGG)
	-10 promoter	-35 promoter	Sequence	Position	
	(TATAAT)	(TTGACA)	(TTTAAGAGGCCAAT)		
gcss					
<i>brajapin8p8gcs</i>	N.D.	-55 to -60	gTTtGcctCCgAT	-21 to -34	-6 to -11
<i>rhietlCIAT652gcs1</i>	N.D.	-44 to -50	N.D.		-7 to -9
<i>rhietlCIAT652gcs2</i>	N.D.	-31 to -36	gTggAGAGGaCcgT	-91 to -104	-2 to -5
<i>rhietl8c3gcs</i>	N.D.	-44 to -51	TTTAAccaGgCAtc	-80 to -93	-5 to -11
<i>rhietlCFN42gcs1</i>	N.D.	-65 to -70	N.D.		N.D.
<i>rhilegGB30gcs1</i>	N.D.	-51 to -55	TgatcGAGGCaAgg	-33 to -46	-8 to -10
<i>rhilegGB30gcs2</i>	N.D.	-47 to -52	gTggtGAGGaCcgT	-90 to -103	-3 to -8
<i>sinfreGR64gcs</i>	N.D.	-23 to -29	TTcAgcggGCCAca	-47 to -60	-6 to -8
<i>sinmel1021gcs</i>	N.D.	-31 to -36	N.D.		-6 to -8
thbs					
<i>azodoeUFLA1-100thb1</i>	N.D.	-54 to -59	TgctgGAcGCCAAc	-95 to -108	-5 to -7
<i>azodoeUFLA1-100thb2</i>	N.D.	-61 to -67	N.D.		-7 to -12
<i>braelkUSDA76thb2</i>	N.D.	-56 to -61	TTTgAGAtaCCtAT	-15 to -28	-3 to -5
<i>braelkUSDA94thb1</i>	N.D.	-34 to -39	gTTgAGAGcGccca	-59 to -72	-2 to -4
<i>brajapUSDA38thb2</i>	N.D.	-38 to -42	TaTAtcAGGgCAca	-23 to -36	-7 to -9
<i>brajapUSDA123thb1</i>	N.D.	-32 to -37	N.D.		-11 to -14
<i>burphySTM815thb1</i>	-8 to -13	-31 to 35	TaTAAacGGtaAcT	-90 to -103	-2 to -4
<i>burphySTM815thb2</i>	-23 to -28	-43 to -47	cTgAtGcGGCCAgc	-70 to -83	N.D.
<i>cupnecN1thb1</i>	N.D.	N.D.	TcgctaAGGCCgcT	47 to -60	-7 to -11
<i>cupnecN1thb2</i>	N.D.	-43 to -47	N.D.		-7 to -9
<i>mescicWSM1271thb</i>	N.D.	-61 to -66	TTgtAGtGGgCgAc	-95 to -108	-8 to -12
<i>meslotNZP2037thb2</i>	N.D.	-39 to -43	TgcAAGccGCCAtc	-47 to -60	-9 to -12
<i>rhietlCNPAF512thb</i>	N.D.	-32 to -37	TaTAtGAGGagcgg	-28 to -41	N.D.
<i>rhietlKIM5thb</i>	N.D.	-59 to -64	N.D.		-8 to -10
<i>rhilegGB30thb2</i>	N.D.	-54 to -58	TTggAatGGaCAAT	-58 to -71	-6 to -10
<i>rhilegVc2thb1</i>	N.D.	-31 to -35	TTcgAcAtGCaAAT	-90 to -103	N.D.
<i>rhilupHPC(L)thb1</i>	-17 to -22	-40 to -45	N.D.		-4 to -7
<i>rhilupHPC(L)thb2</i>	N.D.	N.D.	N.D.		-7 to -11
<i>sinfreHH103thb</i>	N.D.	-42 to -49	TTTgtcAaGCCctg	-102 to -115	-3 to -6 and -8 to -11
<i>sinmel1021thb2</i>	N.D.	-57 to -63	cTTgtcggGCCagAT	-87 to -100	-5 to -7

Sequence alignments and phenetic analysis of rhizobial Glbs

Pairwise sequence alignments showed that the rhizobial fHbs, SDgbs, GCSs and tHbs analyzed in this work are 34.6 to 85.4%, 6.7 to 100%, 10.9 to 100% and 3.5 to 100% identical, respectively. This indicates that variability among the rhizobial Glb sequences is high. Moreover, identity values for the fHbs globin and flavin domains were 39.1 to 93.7% and 26.5 to 81.1%, respectively, and identity values for the GCSs globin and transmitter domains were 17.5 to 100% and 5.9 to 100%, respectively. Thus, apparently in the rhizobial fHbs and GCSs analyzed in this work the globin domain is more conserved than the flavin and transmitter domains.

The average length and molecular mass for the rhizobial fHbs, SDgbs, GCSs and tHbs analyzed in this work are 400 amino acids and 44 kDa, 141 amino acids and 15 kDa, 510 amino acids and 55 kDa and 149 amino acids and 17 kDa, respectively. However, sequence analysis revealed that globin domain from BraelkUSDA76tHb1, BraelkUSDA94tHb1 and Braelk587tHb2 contains 119 to 237 extra amino acids at the N-terminal and 131 extra amino acids at the C-terminal, and that the globin domain from BrajapUSDA123tHb1, BrajapUSDA135tHb1, BraelkWSM1741SDgb2, RhietlCFN42GCS1, BrajapUSDA4tHb2 and BrajapWSM2793tHb3 contains 27 to 73 extra amino acids at the N-terminal. In contrast, a large deletion comprising helices A and B, CD loop and part

A phenogram was constructed from the above multiple sequence alignment. **Figure 2** shows that the rhizobial Glbs analyzed in this work segregate into two main lineages: one containing fHbs, SDgbs and GCSs, and the other containing tHbs (the fHb/SDgb/GCS and tHb lineages, respectively). This is consistent with the main evolutionary lineages identified in bacterial Glbs^{1,51,52} thus indicating that major evolutionary patterns for rhizobial Glbs were identical to those for other bacterial Glbs. Rhizobial fHbs and GCSs cluster with rhizobial SDgbs within the fHb/SDgb/GCS lineage owing to the similarity between the fHb and GCS globin domains and SDgbs. This has been postulated to be the result of an early divergence from a common ancestor to the bacterial fHb and GCS globin domains and SDgbs^{1,6}. The tHb lineage segregates into rhizobial tHbs class 1, tHbs class 2 and tHbs class 3. Within this lineage the rhizobial tHbs class 3 segregate in ancestral position to the rhizobial tHbs class 1 and tHbs class 2. Also, the bradyrhizobial, azorhizobial, mesorhizobial, rhizobial and burkholderial tHbs class 3 segregate from each other; the segregation within rhizobial, sinorhizobial, mesorhizobial and β -rhizobial tHbs class 2 is rather conserved, and bradyrhizobial tHbs class 2 and class 3 segregate into the *B. elkanii* and *B. japonicum* tHb sublineages. These observations indicate that rhizobial tHbs evolved similarly to other bacterial tHbs^{7,8,52} and that evolution of rhizobial tHb sublineages was rather conserved.

Modeling and analysis of the predicted rhizobial Glbs tertiary structure

Structure elucidation is essential to a full understand of a protein's function within the cell's physiology. The structure of a considerable number of bacterial and non-bacterial Glbs has been elucidated by x-ray crystallography. However, with the exception of a *S. meliloti* fHb whose tertiary structure was predicted using bioinformatics methods²⁶, the structure of rhizobial Glbs is not known. Hence, we used bioinformatics methods to predict and analyze the tertiary structure of 44 selected rhizobial Glbs (*i.e.* those representative of major rhizobial Glb clades identified in this work (see **Figure 2** and **Table S2**)) using the best structural homologs as templates (**Dataset 3**).

Predicted structures for selected rhizobial SDgbs and fHbs and GCSs globin domain and tHbs fold into the 3/3- and 2/2-globin fold, respectively (**Figure 3** to **Figure 8**). **Figure 3** shows that structures among the predicted rhizobial fHbs are highly similar. Yet major differences were detected in the BurphySTM815fHb, CupnecHPC(L)fHb and RhilegUMP1137fHb flavin domains, which exhibited two additional helices. **Dataset 3** shows that among globin domains from predicted rhizobial fHbs the distance of the proximal H and distal Q to the heme Fe is 1.44 to 2.47 Å and 6.71 to 15.35 Å, respectively. This observation suggests that the heme Fe in rhizobial fHbs is pentacoordinate.

Figure 4 shows that 3/3-globin folding is highly conserved in the predicted structure of the rhizobial SDgbs AzodoeUFLA1-100SDgb,

BraelkUSDA3254SDgb2, BraelkUSDA3259SDgb1 and BrajapUSDA38SDgb2. Major variations to 3/3-globin folding from predicted rhizobial SDgbs consisted of the existence of an unusually short helix E in BraelkUSDA94SDgb2, a long helix H in BraelkUSDA3254SDgb1 and BrajapUSDA124SDgb1, and the existence of a pre-helix A followed by a long loop at the N-terminal of BraelkWSM1741SDgb2. **Dataset 3** shows that among the predicted rhizobial SDgbs the distance of proximal H and distal Q/R/K/M to the heme Fe is 2.11 to 4.44 Å and 5.08 to 6.63 Å, respectively. This observation suggests that the heme Fe in rhizobial SDgbs is either penta- or hexacoordinate.

Only the globin domain from bacterial GCSs has been crystallized and analyzed by x-ray crystallography^{53,54} (**Dataset 3**). Crystal structure for the bacterial GCSs transmitter domain has not been elucidated. Hence, we only predicted and analyzed the tertiary structure of globin domains from the selected rhizobial GCSs. **Figure 5** shows that the predicted rhizobial GCSs globin domain exhibits a 1.5- to 3-turn pre-helix A, that (with the exception of SinfreGR64GCS) no loop exists between helices A and B, and that helix H is unusually long in Rhietl8C3GCS, RhietlCIAT652GCS2 and RhilegGB30GCS2. **Dataset 3** shows that among the predicted rhizobial GCSs globin domain distance of proximal H/E and distal Q to the heme Fe is 1.77 to 5.56 Å and 4.09 to 9.04 Å, respectively. This observation suggests that the heme Fe in the rhizobial GCSs globin domain is either penta- or hexacoordinate.

Figure 6 to **Figure 8** show that 2/2-globin folding is highly conserved in the predicted rhizobial tHbs class 1, class 2 and class 3. Major variations to 2/2-globin folding from predicted rhizobial tHbs consisted of the existence of a 2.5-turn pre-helix A followed by a long loop at the N-terminal of (class 1) CupnecN1tHb1 (**Figure 6**); the existence of a one-turn pre-helix F (designated as ϕ in **Figure 7**) in the rhizobial tHbs class 2; the existence of a long and extended C-terminal region in (class 2) BraelkUSDA94tHb1 (**Figure 7**), and the substitution of helix A by a long loop that connects to helix B through a 1- to 2.5-turn pre-helix B in (class 3) BraelkUSDA76tHb2, BrajapUSDA123tHb1, BurphySTM815tHb1, Meslot-NZP2037tHb2 and Sinmel1021tHb2 (**Figure 8**). **Dataset 3** shows that among the predicted rhizobial tHbs, the distance of proximal H and distal H/L/F to the heme Fe is 1.77 to 7.51 Å and 4.09 to 8.25 Å, respectively. This observation suggests that the heme Fe in the rhizobial tHbs is either penta- or hexacoordinate.

The above observations suggest that in spite of sequence variability (see the *Sequence alignments and phenetic analysis of rhizobial Glbs* subsection) the structure of rhizobial Glbs is similar to the canonical 3/3- or 2/2-globin folding of bacterial and non-bacterial Glbs. However, a number of predicted rhizobial Glbs exhibited variations at the N- and C-terminal regions suggesting that their structural properties could be different to those of canonical Glbs.

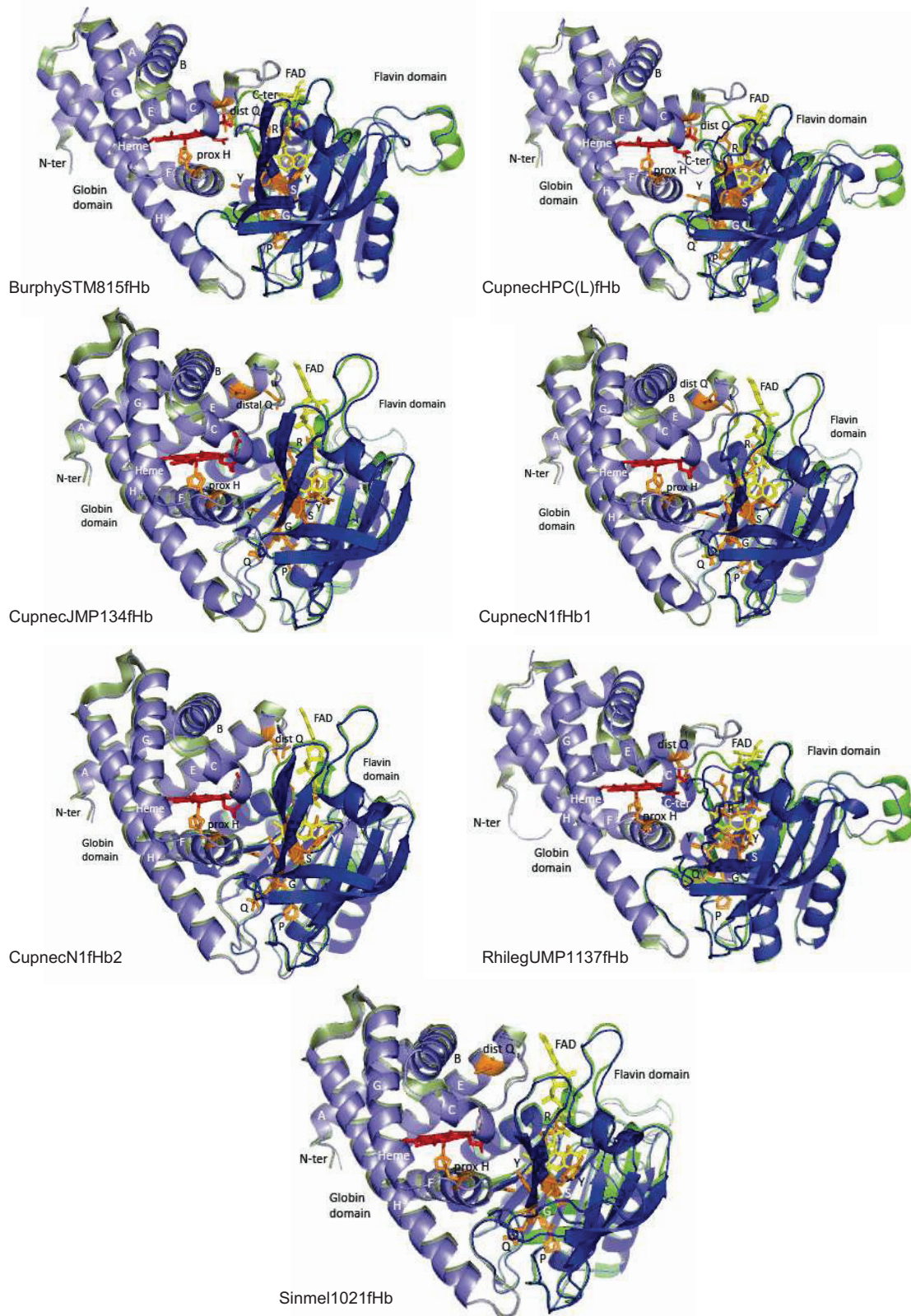


Figure 3. Predicted structure of rhizobial fHbs (blue) overlapped to structural homologues (green). Structural homologues are indicated in Dataset 3. Distal and proximal amino acids to the heme Fe and amino acids that interact with the FAD cofactor are shown in brown. Heme and FAD are shown in red and yellow, respectively. Helices within the globin domain are indicated with letters A to H. All structures are displayed in the same orientation.

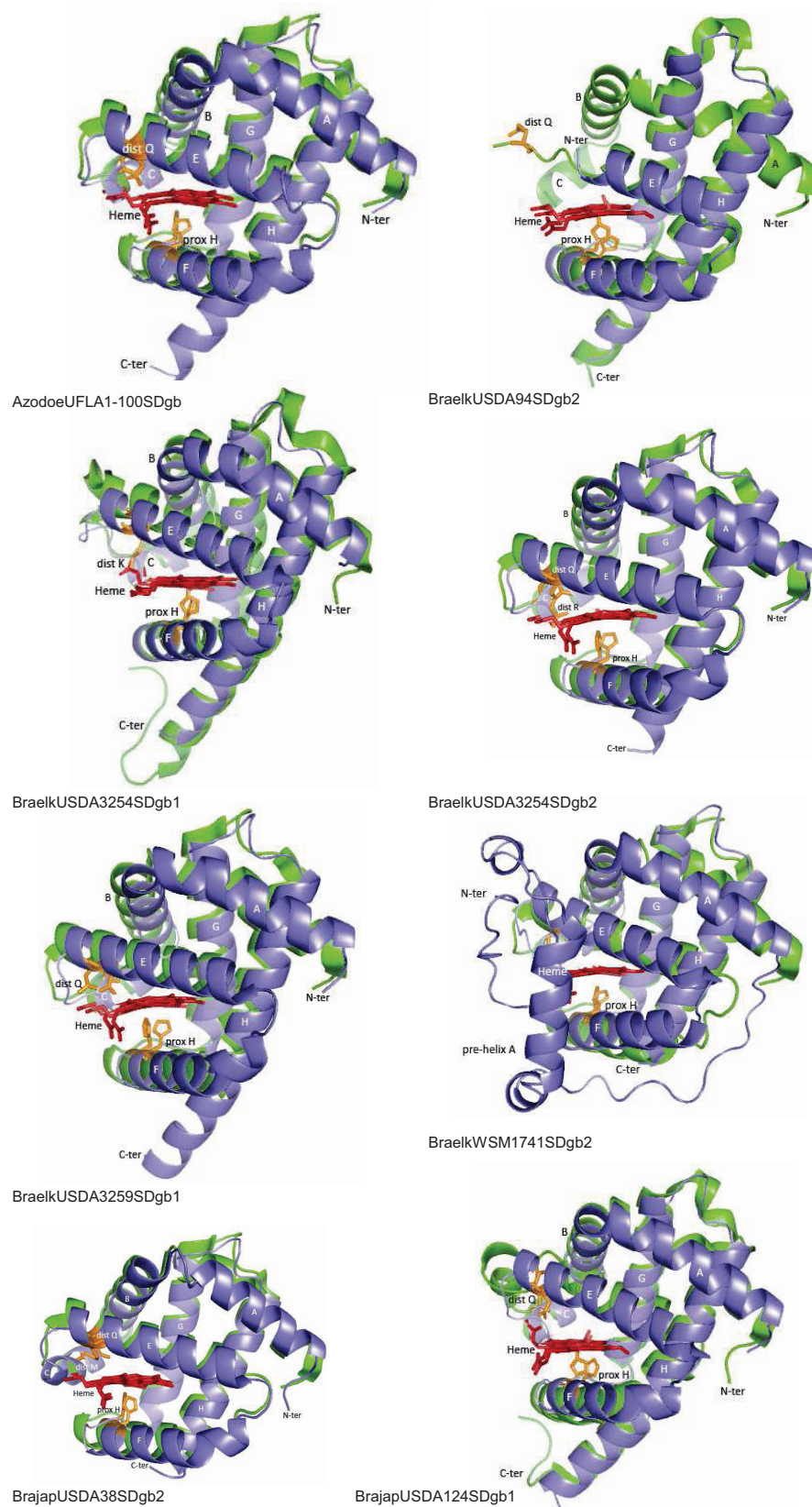
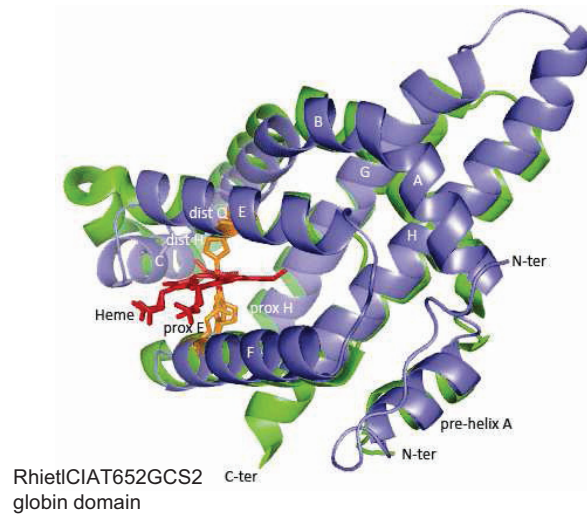
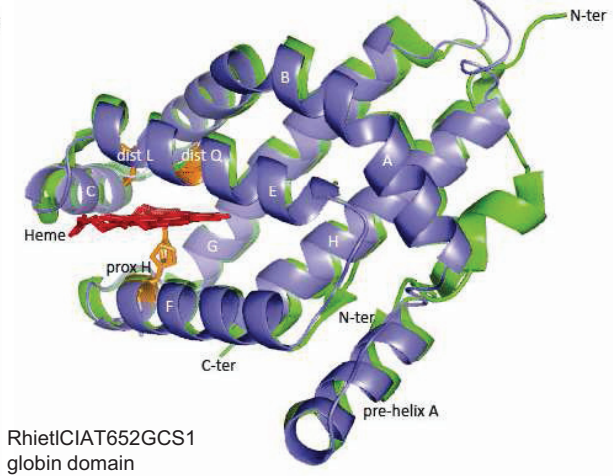
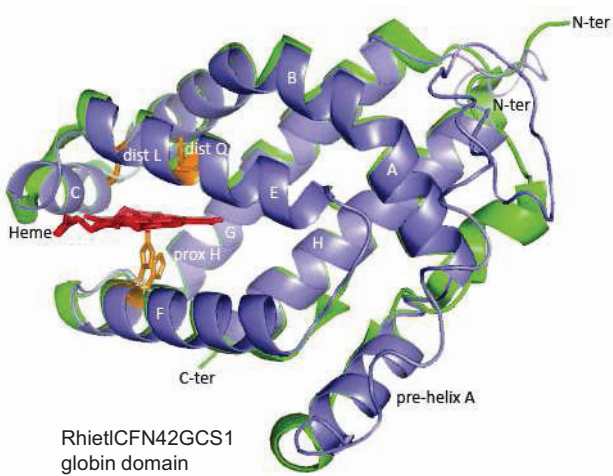
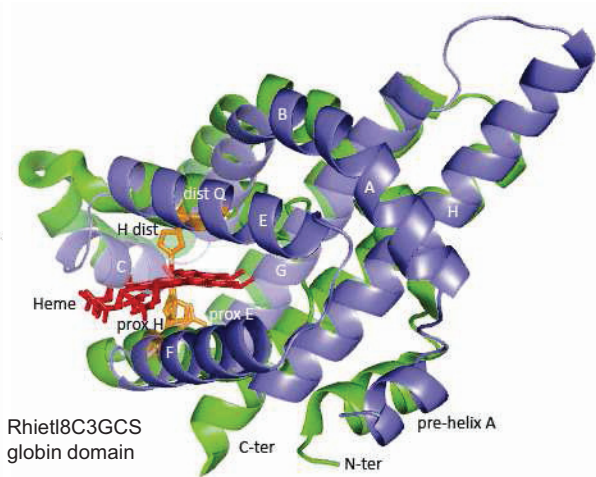
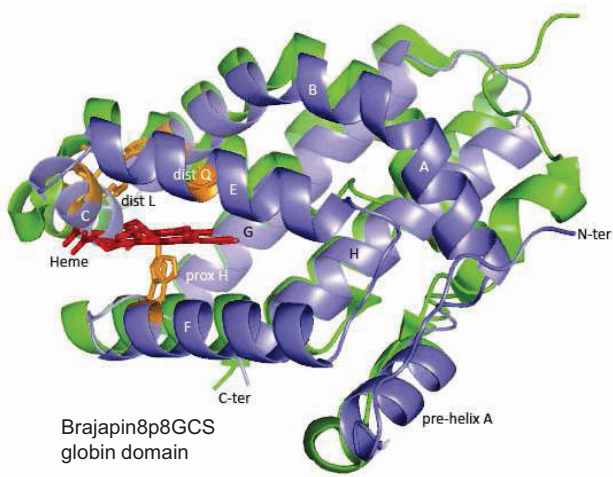


Figure 4. Predicted structure of selected rhizobial SDgbs (blue) overlapped to structural homologues (green). Structural homologues are indicated in Dataset 3. Distal and proximal amino acids to the heme Fe are shown in brown. Heme is shown in red. Helices are indicated with letters A to H. All structures are displayed in the same orientation.



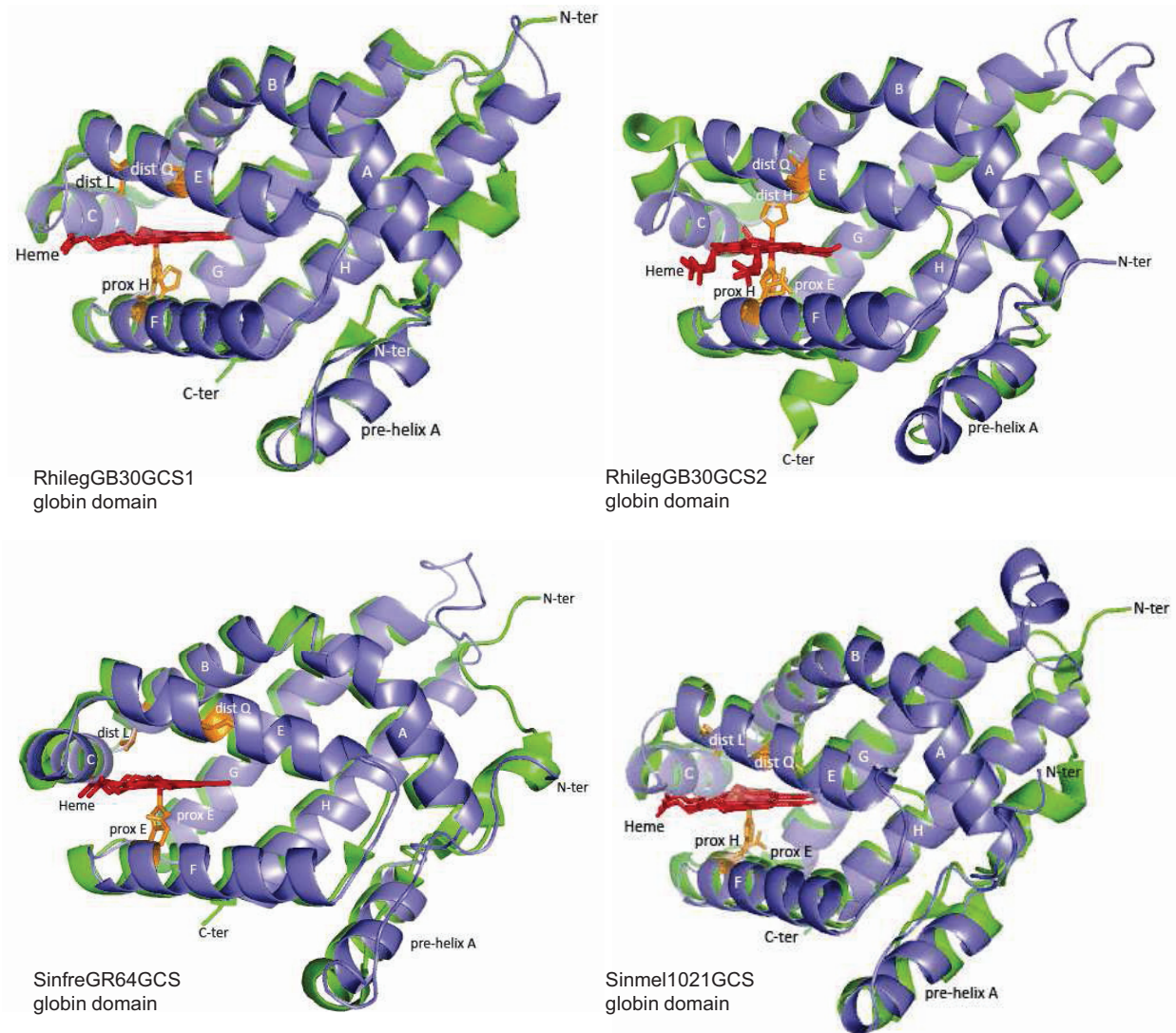


Figure 5. Predicted structure of selected rhizobial GCSs globin domain (blue) overlapped to structural homologues (green). Structural homologues are indicated in Dataset 3. Distal and proximal amino acids to the heme Fe are shown in brown. Heme is shown in red. Helices are indicated with letters A to H. All structures are displayed in the same orientation.

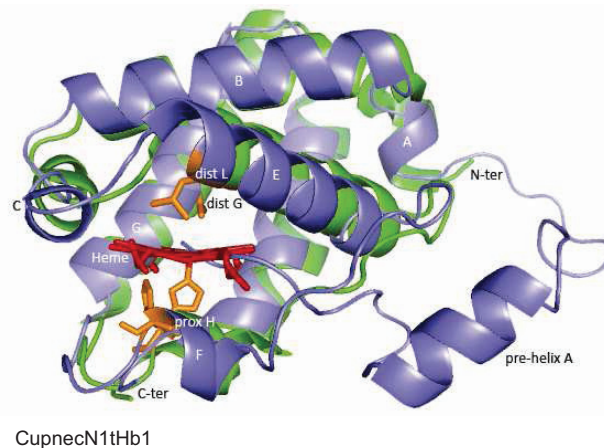
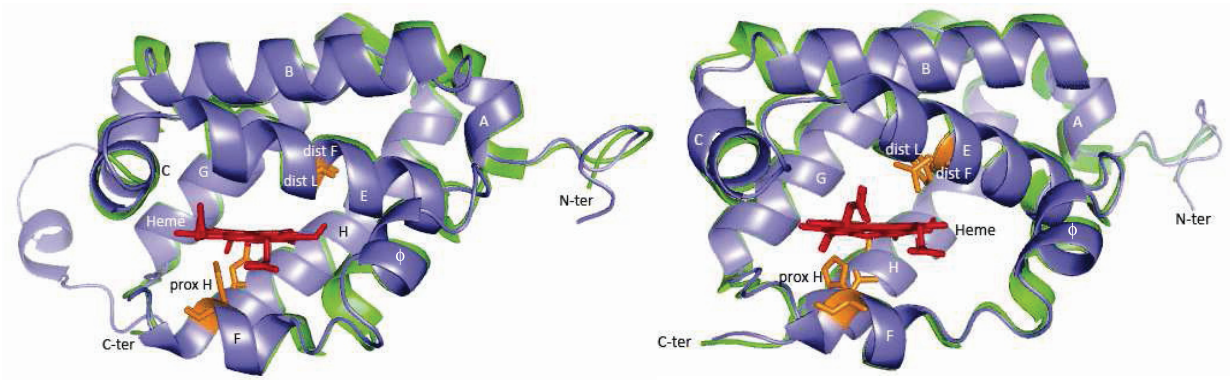
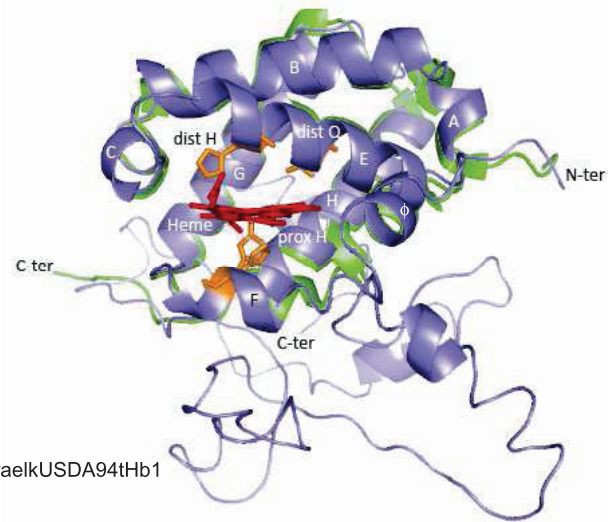


Figure 6. Predicted structure of class 1 CupnecN1tHb1 (blue) overlapped to the structural homologue *Tetrahymena pyriformis* tHb (PDB ID 3AQ5) (green). Distal and proximal amino acids to the heme Fe are shown in brown; only potential distal E11 is shown in the CupnecN1tHb1 structure. Heme is shown in red. Helices are indicated with letters A to H.

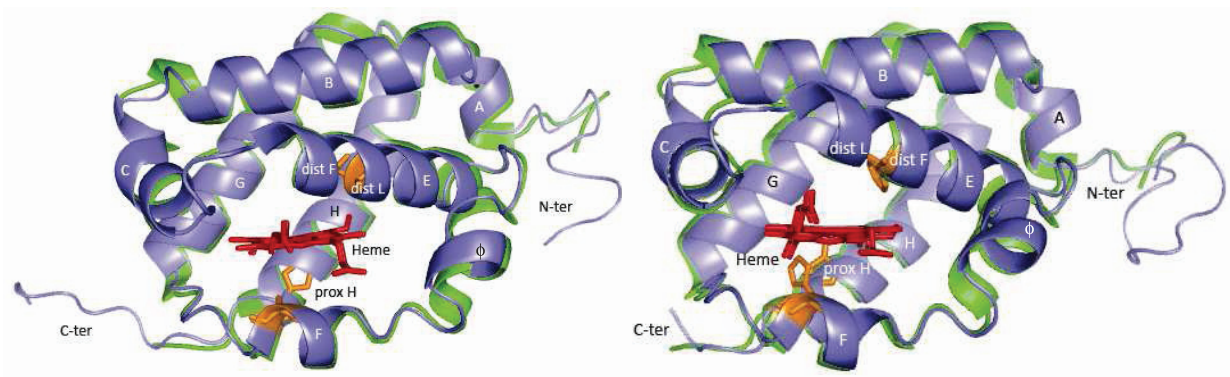


AzodoeUFLA1-100tHb2

BrajapUSDA38tHb2



BraelkUSDA94tHb1



BurphySTM815tHb2

CupnecN1tHb2

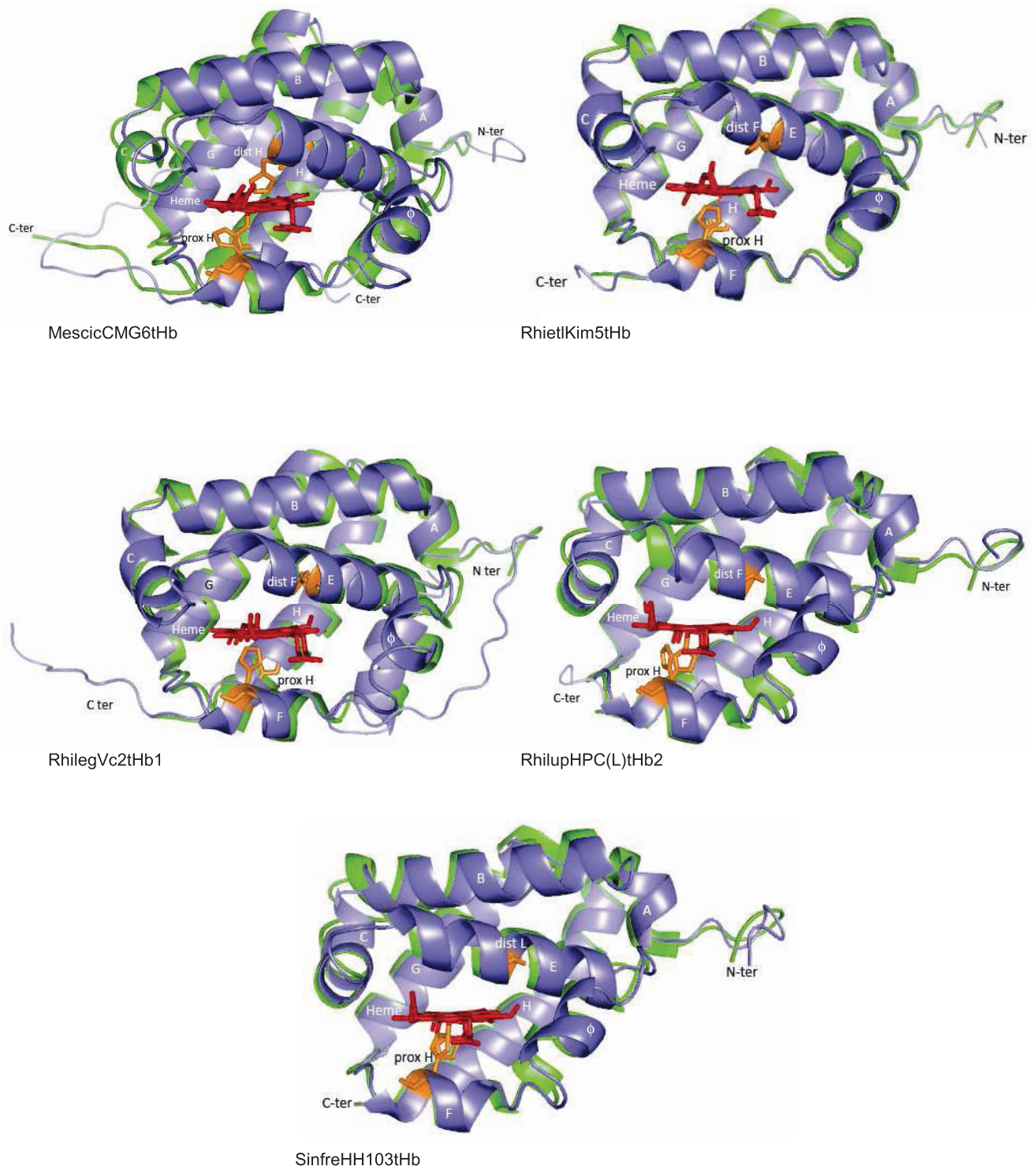
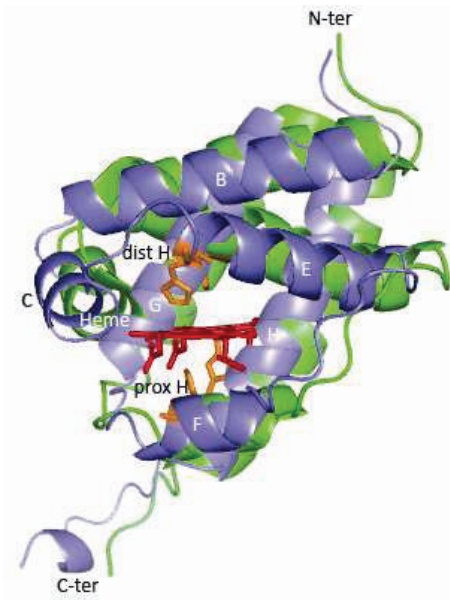
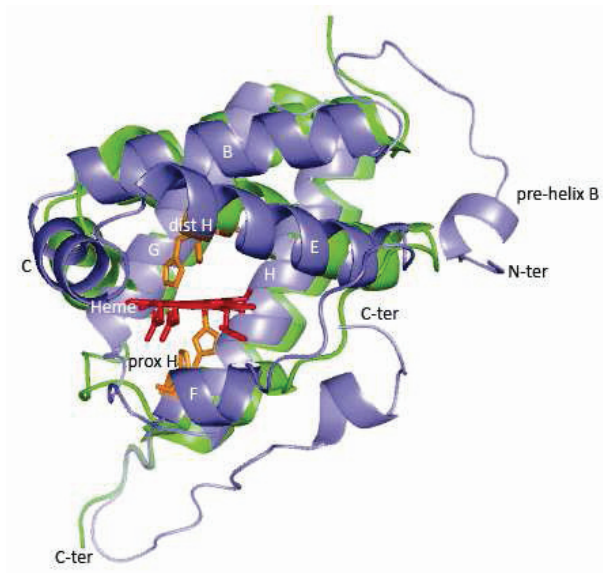


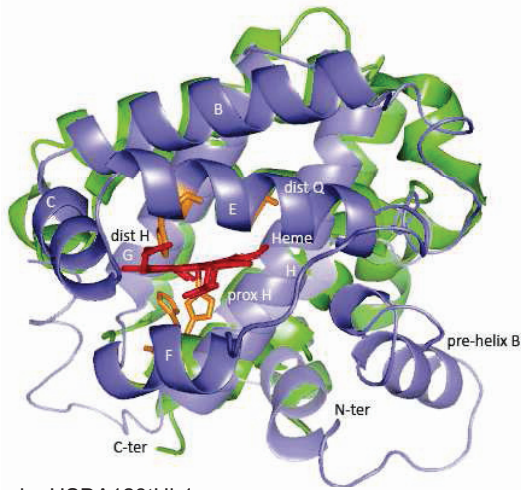
Figure 7. Predicted structure of selected rhizobial tHbs class 2 (blue) overlapped to structural homologues (green). Structural homologues are indicated in [Dataset 3](#). Distal and proximal amino acids to the heme Fe are shown in brown; only potential distal E11 is shown in the tHbs structure. Heme is shown in red. Helices are indicated with letters A to H. Pre-helix F is indicated with the Greek letter ϕ . All structures are displayed in the same orientation.



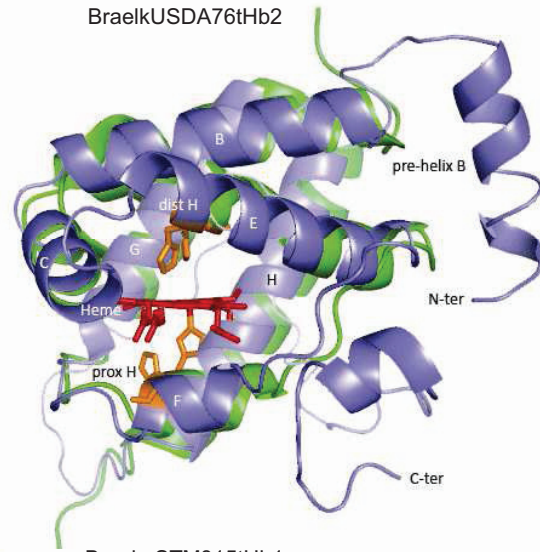
AzodoeUFLA1-100tHb1



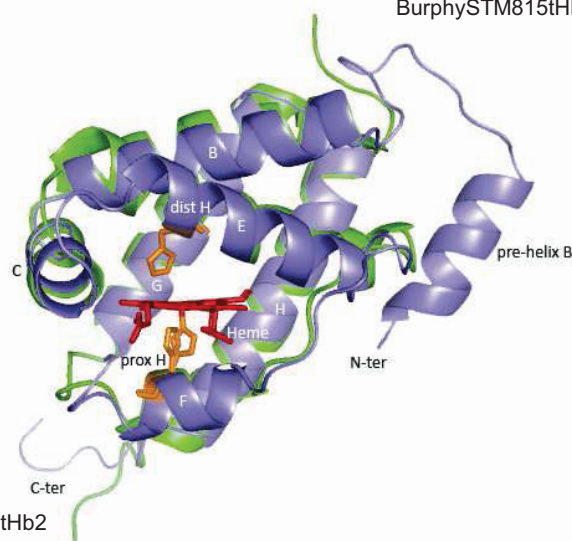
BraelkUSDA76tHb2



BrajapUSDA123tHb1



BurphySTM815tHb1



MeslotNZP2037tHb2

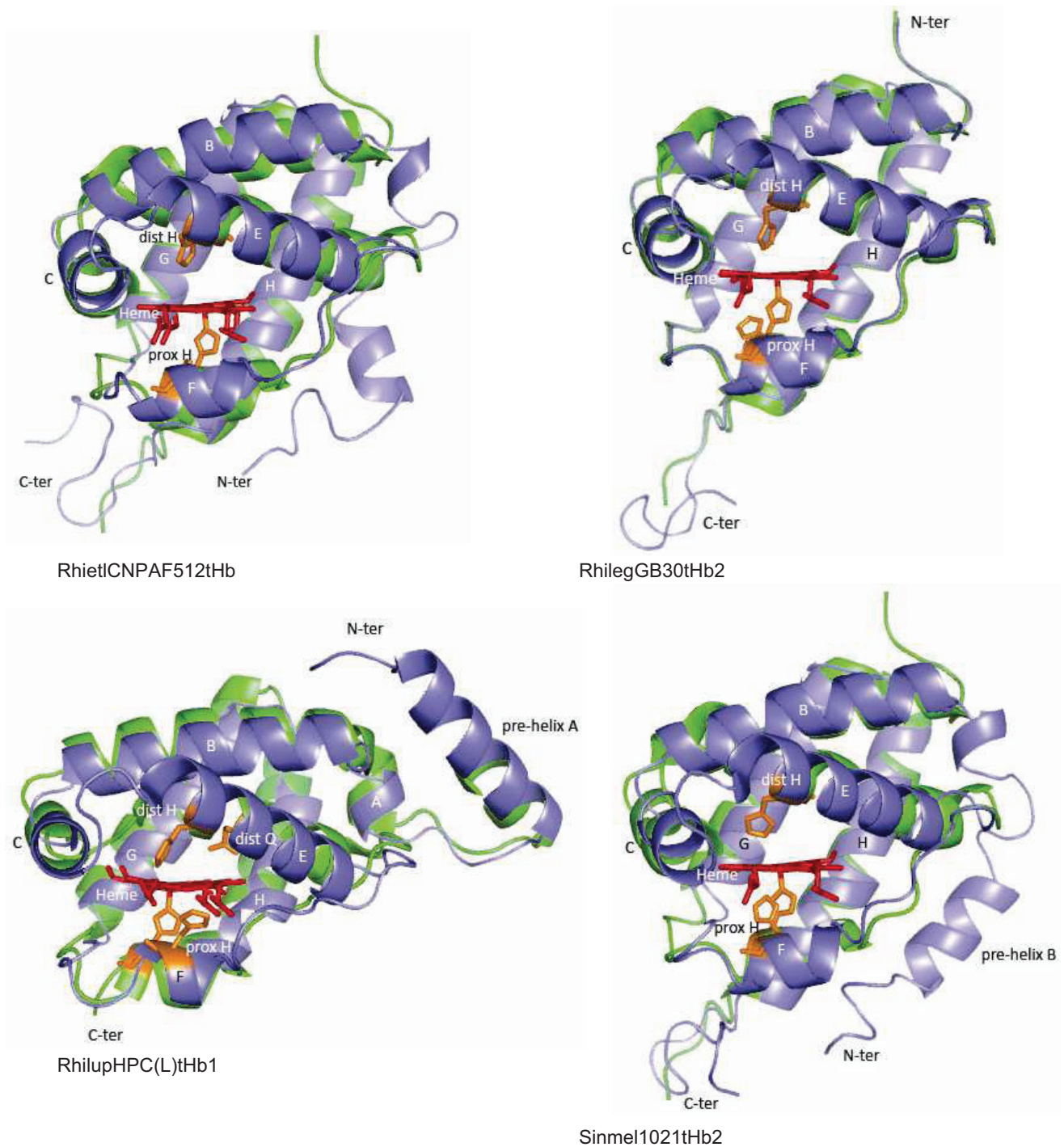


Figure 8. Predicted structure of selected rhizobial tHbs class 3 (blue) overlapped to structural homologues (green). Structural homologues are indicated in [Dataset 3](#). Distal and proximal amino acids to the heme Fe are shown in brown; only potential distal E11 is shown in the tHbs structure. Heme is shown in red. Helices are indicated with letters A to H. All structures are displayed in the same orientation.

Data also shows that (with few exceptions) in addition to proximal and distal amino acids the distance of B10 and CD1 amino acids to the heme Fe and the orientation of proximal, distal, B10 and CD1 amino acids are similar within and among the predicted rhizobial SDgbs, fHbs and GCSs globin domain and tHbs. These amino acids participate in the binding of ligands to the heme Fe. Thus, these observations suggest that the mechanisms and chemistry for ligand binding are similar among the rhizobial Glbs.

Spectroscopic identification of putative Glbs in soluble extracts from *Bradyrhizobium japonicum* USDA38 and USDA58

The prerequisites for being able to infer a protein's function are isolating and characterizing either native or recombinant proteins and detecting protein synthesis *in vivo*. No rhizobial Glb has been isolated and characterized thus far. However, spectroscopic evidence indicates that putative Glbs exist in soluble extracts from *B. japonicum* 505 (Wisconsin), *R. leguminosarum* bv. *viciae*, *B. japonicum* NPK63 and *R. etli* CE3 (see the *Introduction* section). In order to extend these analyses to other rhizobia, we

analyzed soluble extracts from *B. japonicum* USDA38 and USDA58 by (dithionite reduced + CO minus dithionite reduced) differential spectroscopy using as controls the sperm whale myoglobin and bovine blood hemoglobin. **Table 3** shows that absorption peaks and troughs in the Soret and Q regions for the *B. japonicum* USDA38 and USDA58, *B. japonicum* 505 (Wisconsin), *R. leguminosarum* bv. *viciae*, *B. japonicum* NPK63 and *R. etli* CE3 soluble extracts, *Vitreoscilla* VHb, *E. coli* K12 Hmp, sperm whale myoglobin and bovine blood hemoglobin are nearly identical. This preliminary evidence indicates that putative soluble Glbs are synthesized in *B. japonicum* USDA38 and USDA58. Interestingly, genes coding for SDgbs (*brajapUSDA38SDgb1* and *brajapUSDA38SDgb2*) and tHbs (*brajapUSDA38tHb1* and *brajapUSDA38tHb2*) were identified in the *B. japonicum* USDA38 genome (**Dataset 1**). Thus, it is likely that putative *B. japonicum* USDA38 Glbs corresponds to a combination of SDgbs and tHbs. Inferences from the preliminary results reported here should be confirmed by Glb detection, isolation and unequivocal identification after protein sequencing. This may open the possibility to carry out further experimental analyses on rhizobial Glbs.

Table 3. Absorption peaks and troughs in the Soret and Q regions from the (dithionite reduced + CO minus dithionite reduced) differential spectra of rhizobial soluble extracts and other bacterial and vertebrate Glbs.

Rhizobial soluble extract/Glb	Soret region		Q region				Reference
	Peak (nm)	Trough (nm)	Peak (nm)	Trough (nm)	Peak (nm)	Trough (nm)	
Rhizobial soluble extracts							
<i>B. japonicum</i> USDA38	425	448	535	573	549	600	This work
<i>B. japonicum</i> USDA58	416	437	535	573	554	601	This work
<i>B. japonicum</i> NPK63	422	443	529	574	558	598	24
<i>B. japonicum</i> 505 (Wisconsin)	417	434	540	569	556	n.i.	22
<i>R. etli</i> CE3	421	439	539	563	547	590	25
<i>R. leguminosarum</i> bv. <i>viciae</i> 96	424	443	535	574	555	n.i.	23
Bacterial Glbs							
<i>Vitreoscilla</i> VHb	418	436	534	567	551	590	25
<i>E. coli</i> K12 Hmp	420	437	530	570	555	592	55
Vertebrate Glbs							
Sperm whale myoglobin	419	436	538	578	558	596	This work
Bovine blood hemoglobin	417	432	533	570	554	588	This work

n.i., non-identified

Conclusions

Rhizobial Glbs have been poorly studied. However, results reported in this work provide molecular and biochemical data from a bioinformatics perspective that contribute to a better understanding of these proteins. For example, the distribution and outline for the evolution of *glb* genes and Glb proteins among rhizobia was clarified, genes that could coexpress with the rhizobial *glbs* were identified and the predicted tertiary structure for rhizobial Glbs was elucidated. Also, spectroscopic analysis suggested that soluble Glbs are synthesized in free-living *B. japonicum* USDA38 and USDA58. This information will be useful in designing future experimental work focused on clarifying Glb functions within the physiology of free-living and symbiotic rhizobia.

Data availability

F1000Research: Dataset 1. Globin genes detected in the genomes of rhizobial bacteria. [10.5256/f1000research.6392.d46189](https://doi.org/10.5256/f1000research.6392.d46189)

F1000Research: Dataset 2. Predicted Glb polypeptides detected in the genomes of rhizobial bacteria. [10.5256/f1000research.6392.d46190](https://doi.org/10.5256/f1000research.6392.d46190)

F1000Research: Dataset 3. Distance to the heme Fe and orientation of distal, proximal, B10 and CD1 amino acids in the predicted structure of selected rhizobial Glbs (Table S2). [10.5256/f1000research.6392.d46191](https://doi.org/10.5256/f1000research.6392.d46191)

Author contributions

RGB, MSS and RAP conceived the study. RGB and MSS executed the experiments. RAP prepared the first draft of the manuscript. RGB, MSS and RAP revised the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

This work was partially financed by SEP-PROMEP (grant number UAEMor-PTC-01-01/PTC23) and Consejo Nacional de Ciencia y Tecnología (CoNaCyT grant numbers 25229N and 42873Q), México. R. Gesto-Borroto is a graduate student financially supported by CoNaCyT (registration no. 293307).

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Drs. Donald Keister and Douglas Jones (United States Department of Agriculture, USA) for kindly providing the *Bradyrhizobium japonicum* USDA38 and USDA58 strains, and Dr. Serge N. Vinogradov (Wayne State University, Detroit MI, USA) for critical reading of this article and providing constructive comments.

Supplementary materials

Supplementary File S1. Mapping of the *fhb* (A), *sdgb* (B), *gcs* (C) and *thb* (D) genes in the genomes of rhizobial bacteria. DNA fragments correspond to ~5 kb up- and downstream to *glb* genes. Arrows indicate the transcription orientation. The *glb* genes are shown in red, predicted polypeptides functioning in nitrogen metabolism are shown in blue and predicted polypeptides functioning in chemotaxis are shown in green. The ORF sizes and distances between ORFs are shown at an approximate scale. Abbreviations for the predicted polypeptides are indicated at the end of the figure. Location of DNA fragments in the rhizobial genome: Chr, chromosome; Pmd, plasmid; n.i., non-identified.

Supplementary File S2. Sequence alignment of Glbs detected in the genomes of rhizobial bacteria. The tHb, fHb, SDgb and GCS sequences are shown in green, light blue, dark blue and red, respectively. Distal and proximal amino acids located in helices E and F, respectively, are indicated within black boxes; potential distal E7 and E11 are indicated in the tHb sequences. Amino acids that interact with FAD and NAD(P)⁺ cofactors in the fHb flavin domain are indicated within gray boxes. Limits for the globin domain are indicated with right- and left-oriented arrows within black circles. Helices are indicated with letters A to H within the 2/2- and 3/3-fold of the tHb and SDgb and fHb and GCS globin domains, respectively. Outgroups for fHbs correspond to BacsubfHb, EsccolfHb and SaccerrfHb (Genbank accession number YP_003865693, NP_289108 and NP_011750, respectively); outgroup for SDgbs corresponds to VitSDgb (Genbank accession number AAA75506); outgroups for GCSs correspond to AgrtumGCS and BacsubGCS (Genbank accession number NP_354049 and NP_388919, respectively); outgroups for tHbs correspond to MyctubtHb class1, MyctubtHb class 2, AgrtumtHb class 2 and MycavitHb class 3 (Genbank accession number NP_216058, NP_216986, WP_020813663 and BAN32501, respectively).

Table S1. Classification of the α - and β -rhizobia and rhizobial genomes analyzed in this work.

α-rhizobia			
Domain Eubacteria Phylum Proteobacteria Class Alphaproteobacteria Order Rhizobiales			
Family	Genus	Species/strain	Database
Bradyrhizobiaceae	<i>Bradyrhizobium</i>	<i>B. elkanii</i> USDA3254, 587, USDA3259, USDA76, USDA94, WSM1741, WSM2783 <i>B. japonicum</i> USDA110, 22, USDA122, USDA123, USDA124, USDA 135, USDA38, USDA4, USDA6-7488, USDA6-8372, WSM1743, WSM2793, in8p8, is5	JGI Genome Portal (http://genome.jgi.doe.gov/) Rhizobase (http://genome.microbedb.jp/rhizobase)
Phyllobacteriaceae	<i>Mesorhizobium</i>	<i>M. ciceri</i> CMG6, WSM4083, bv. biserrulae WSM1271 <i>M. loti</i> CJ3sym, MAFF303099, NZP2037, R7A, R88b, USDA3471	JGI Genome Portal (http://genome.jgi.doe.gov/) Rhizobase (http://genome.microbedb.jp/rhizobase)
Rhizobiaceae	<i>Rhizobium</i>	<i>R. etli</i> 8C-3, Brasil 5, CFN42, CIAT652, CIAT894, CNPAF512, GR56, IE4771, Kim5 <i>R. leguminosarum</i> bv. viciae UPM1131, bv. viciae UPM1137, bv. viciae 128C53, bv. viciae GB30, bv. viciae VF39, bv. viciae WSM1455, bv. viciae WSM1481, bv. viciae 248, bv. viciae 3841, bv. viciae Ps8, bv. viciae TOM, bv. viciae Vh3, bv. viciae Vc2 <i>R. lupini</i> HPC(L)	CCG-UNAM (http://www.cifn.unam.mx/) Sanger Institute (http://www.sanger.ac.uk) NEERI (http://www.neeri.res.in/)
	<i>Sinorhizobium</i>	<i>S. fredii</i> GR64, HH103, USDA257 <i>S. meliloti</i> 1021	GGL (http://appmibio.uni-goettingen.de/) INRA (https://iant.toulouse.inra.fr/)
Xanthobacteraceae	<i>Azorhizobium</i>	<i>A. doebereinae</i> UFLA1-100	JGI Genome Portal (http://genome.jgi.doe.gov/)
β-rhizobia			
Domain Eubacteria Phylum Proteobacteria Class Betaproteobacteria Order Burkholderiales			
Family	Genus	Species/strain	Database
Burkholderiaceae	<i>Burkholderia</i>	<i>B. phymatum</i> STM815	JGI Genome Portal (http://genome.jgi.doe.gov/)
	<i>Cupriavidus</i>	<i>C. necator</i> * HPC(L), JMP134, N-1 ATCC43291	Goettingen Genomics Laboratory (http://appmibio.uni-goettingen.de/)

*Formerly classified as *Alcaligenes eutrophus* and *Ralstonia eutropha*.

Table S2. Predicted structures of rhizobial Glbs deposited in the Caspur protein model dataBase (<http://bioinformatics.cineca.it/PMDB/>).

Glb	ID number
fHbs	
BurphySTM815fHb	PM0079658
CupnecHPC(L)fHb	PM0079659
CupnecJMP134fHb	PM0079660
CupnecN-1 ATCC43291fHb1	PM0079661
CupnecN-1 ATCC43291fHb2	PM0079662
RhilegUPM1137fHb	PM0079663
Sinmel1021fHb	PM0079672
SDgbs	
AzodoeUFLA1-100SDgb	PM0079664
BraelkUSDA94SDgb2	PM0079665
BraelkUSDA3254SDgb1	PM0079666
BraelkUSDA3254SDgb2	PM0079667
BraelkUSDA3259SDgb1	PM0079668
BraelkWSM1741SDgb2	PM0079669
BrajapUSDA38SDgb2	PM0079670
BrajapUSDA124SDgb1	PM0079671
GCSs	
Brajapin8p8GCS globin domain	PM0079673
RhietICIAT652GCS1 globin domain	PM0079674
RhietICIAT652GCS2 globin domain	PM0079675
Rhietl8C-3GCS globin domain	PM0079676
RhietICFN42DSM 11541GCS1 globin domain	PM0079677

Glb	ID number
RhilegGB30GCS1 globin domain	PM0079678
RhilegGB30GCS2 globin domain	PM0079679
SinfreGR64GCS globin domain	PM0079680
Sinmel1021GCS globin domain	PM0079681
tHbs	
AzodoeUFLA1-100tHb1	PM0079682
AzodoeUFLA1-100tHb2	PM0079683
BraelkUSDA76tHb2	PM0079684
BraelkUSDA94tHb1 (globin domain + C-terminal extension)	PM0079701
BrajapUSDA38tHb2	PM0079685
BrajapUSDA123tHb1	PM0079686
BurphySTM815tHb1	PM0079687
BurphySTM815tHb2	PM0079688
CupnecN-1 ATCC43291tHb1	PM0079689
CupnecN-1 ATCC43291tHb2	PM0079690
MescicCMG6tHb	PM0079691
MeslotNZP2037tHb2	PM0079692
RhietICNPAF512tHb	PM0079693
RhietlKim5tHb	PM0079694
RhilegGB30tHb2	PM0079695
RhilegVc2tHb1	PM0079696
RhilupHPC(L)tHb1	PM0079697
RhilupHPC(L)tHb2	PM0079698
SinfreHH103tHb	PM0079699
Sinmel1021tHb2	PM0079700

References

1. Vinogradov SN, Hoogewijs D, Bailly X, *et al.*: **A phylogenomic profile of globins.** *BMC Evol Biol.* 2006; **6**: 31–47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Dickerson RE, Geis I: **Hemoglobin: structure, function, evolution, and pathology.** Menlo Park, California: The Benjamin/Cummings Pub. Co., Inc.; 1983; 176.
[Reference Source](#)
3. Nardini M, Pesce A, Milani M, *et al.*: **Protein fold and structure in the truncated (2/2) globin family.** *Gene.* 2007; **398**(1–2): 2–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Hardison R: **The evolution of hemoglobin.** *Am Sci.* 1999; **87**(2): 126–137.
[Publisher Full Text](#)
5. Wajcman H, Kiger L: **Hemoglobin, from microorganisms to man: a single structural motif, multiple functions.** *C R Biol.* 2002; **325**(12): 1159–1174.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Vinogradov SN, Hoogewijs D, Bailly X, *et al.*: **Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life.** *Proc Natl Acad Sci USA.* 2005; **102**(32): 11385–11389.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Vinogradov SN, Tinajero-Trejo M, Poole RK, *et al.*: **Bacterial and archaeal globins - A revised perspective.** *Biochim Biophys Acta.* 2013; **1834**(9): 1789–1800.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Vuletic DA, Lecomte JT: **A phylogenetic and structural analysis of truncated hemoglobins.** *J Mol Evol.* 2006; **62**(2): 196–210.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Wittenberg JB, Bolognesi M, Wittenberg BA, *et al.*: **Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants.** *J Biol Chem.* 2002; **277**(2): 871–874.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Gardner PR: **Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases.** *J Inorg Biochem.* 2005; **99**(1): 247–266.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Gardner PR, Gardner AM, Brashear WT, *et al.*: **Hemoglobins dioxygenate nitric oxide with high fidelity.** *J Inorg Biochem.* 2006; **100**(4): 542–550.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Giardina B, Messana I, Scatena R, *et al.*: **The multiple functions of hemoglobin.** *Crit Rev Biochem Mol Biol.* 1995; **30**(3): 165–196.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Vinogradov SN, Moens L: **Diversity of globin function: enzymatic, transport, storage, and sensing.** *J Biol Chem.* 2008; **283**(14): 8773–8777.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Mylona P, Pawlowski K, Bisseling T: **Symbiotic nitrogen fixation.** *Plant Cell.* 1995; **7**(7): 869–885.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Agron PG, Ditta GS, Helinski DR: **Oxygen regulation of nifA transcription in vitro.** *Proc Natl Acad Sci USA.* 1993; **90**(8): 3506–3510.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Boscarri A, Meilhoc E, Castella C, *et al.*: **Which role for nitric oxide in symbiotic N₂-fixing nodules: toxic by-product or useful signaling/metabolic intermediate?** *Front Plant Sci.* 2013; **4**: 384.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. del-Giudice J, Cam Y, Damiani I, *et al.*: **Nitric oxide is required for an optimal establishment of the Medicago truncatula-Sinorhizobium meliloti symbiosis.** *New Phytol.* 2011; **191**(2): 405–417.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Appleby CA: **The origin and functions of haemoglobin in plants.** *Sci Progress.* 1992; **76**: 365–398.
[Reference Source](#)
19. Downie JA: **Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules.** *Curr Biol.* 2005; **15**(6): R196–198.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Herold S, Puppo A: **Kinetics and mechanistic studies of the reactions of metleghemoglobin, ferryleghemoglobin, and nitrosylleghemoglobin with reactive nitrogen species.** *J Biol Inorg Chem.* 2005; **10**(8): 946–957.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Herold S, Puppo A: **Oxyleghemoglobin scavenges nitrogen monoxide and peroxytrite: a possible role in functioning nodules?** *J Biol Inorg Chem.* 2005; **10**(8): 935–945.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Appleby CA: **Electron transport systems of Rhizobium japonicum. II. Rhizobium, haemoglobin cytochromes and oxidases in free-living (cultured) cells.** *Biochim Biophys Acta.* 1969; **172**(1): 88–105.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Kretovich WL, Romanov VI, Korolyov AV: **Rhizobium leguminosarum cytochromes (Vicia faba).** *Plant and Soil.* 1973; **39**(3): 619–634.
[Publisher Full Text](#)
24. Keister DL, Marsh SS: **Hemoproteins of Bradyrhizobium japonicum cultured cells and bacteroids.** *Appl Environ Microbiol.* 1990; **56**(9): 2736–2741.
[PubMed Abstract](#) | [Free Full Text](#)
25. Ramírez M, Valderrama B, Arredondo-Peter R, *et al.*: **Rhizobium etli genetically engineered for the heterologous expression of Vitreoscilla sp. hemoglobin: effects on free-living and symbiosis.** *Mol Plant-Microbe Interact.* 1999; **12**(11): 1008–1015.
[Publisher Full Text](#)
26. Lira-Ruan V, Sarath G, Klucas RV, *et al.*: **In silico analysis of a flavohemoglobin from Sinorhizobium meliloti strain 1021.** *Microbiol Res.* 2003; **158**(3): 215–227.
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Meilhoc E, Cam Y, Skapski A, *et al.*: **The response to nitric oxide of the nitrogen-fixing symbiont Sinorhizobium meliloti.** *Mol Plant-Microbe Interact.* 2010; **23**(6): 748–759.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Gough J, Karplus K, Hughey R, *et al.*: **Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure.** *J Mol Biol.* 2001; **313**(4): 903–919.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Shi J, Blundell TL, Mizuguchi K: **FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties.** *J Mol Biol.* 2001; **310**(1): 243–257.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Kiley PJ, Beinert H: **Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster.** *FEMS Microbiol Rev.* 1998; **22**(5): 341–352.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Thompson JD, Gibson TJ, Plewniak F, *et al.*: **The clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucl Acids Res.* 1997; **25**(24): 4876–4882.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Kapp OH, Moens L, Vanfleteren J, *et al.*: **Alignment of 700 globin sequences: extent of amino acid substitution and its correlation with variation in volume.** *Prot Sci.* 1995; **4**(10): 2179–2190.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Lesk AM, Chothia C: **How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins.** *J Mol Biol.* 1980; **136**(3): 225–270.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Letunic I, Bork P: **Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.** *Nucl Acids Res.* 2011; **39**(Web Server issue): W475–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nature Protoc.* 2010; **5**(4): 725–738.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Roy A, Xu D, Poisson J, *et al.*: **A protocol for computer-based protein structure and function prediction.** *J Vis Exp.* 2011; (57): e3259.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics.* 2008; **9**: 40.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Humphrey W, Dalke A, Schulten K: **VMD: Visual molecular dynamics.** *J Mol Graph.* 1996; **14**(1): 33–38, 27–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Gopalasubramanian SK, Garrocho-Villegas V, Rivera GB, *et al.*: **Use of in silico (computer) methods to predict and analyze the tertiary structure of plant hemoglobins.** *Meth Enzymol.* 2008; **436**: 393–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Sáenz-Rivera J, Sarath G, Arredondo-Peter R: **Modeling the tertiary structure of a maize (Zea mays ssp. mays) non-symbiotic hemoglobin.** *Plant Physiol Biochem.* 2004; **42**(11): 891–897.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Cruz-Ramos H, Crack J, Wu G, *et al.*: **NO sensing by FNR: regulation of the Escherichia coli NO-detoxifying flavohemoglobin, Hmp.** *EMBO J.* 2002; **21**(13): 3235–3244.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Joshi M, Dikshit KL: **Oxygen dependent regulation of Vitreoscilla globin gene: evidence for positive regulation by FNR.** *Biochem Biophys Res Comm.* 1994; **202**(1): 535–542.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Khroshilova N, Pepescu C, Munck E, *et al.*: **Iron-sulfur cluster disassembly in the FNR protein of Escherichia coli by O₂: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity.** *Proc Natl Acad Sci USA.* 1997; **94**(12): 6087–6092.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Poole RK, Anjum MF, Membrillo-Hernández J, *et al.*: **Nitric oxide, nitrite, and Fnr regulation of hmp (flavohemoglobin) gene expression in Escherichia coli K-12.** *J Bacteriol.* 1996; **178**(18): 5487–5492.
[PubMed Abstract](#) | [Free Full Text](#)

45. Vinogradov SN, Fernández I, Hoogewijs D, *et al.*: **Phylogenetic relationships of 3/3 and 2/2 hemoglobins in Archaeplastida genomes to bacterial and other eukaryote hemoglobins.** *Mol Plant.* 2011; **4**(1): 42–58.
[PubMed Abstract](#) | [Publisher Full Text](#)
46. Vinogradov SN, Waltz DA, Pohajdak B, *et al.*: **Adventitious variability? The amino acid sequences of nonvertebrate globins.** *Comp Biochem Physiol B.* 1993; **106**(1): 1–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Weber RE, Vinogradov SN: **Nonvertebrate hemoglobins: functions and molecular adaptations.** *Physiol Rev.* 2001; **81**(2): 569–628.
[PubMed Abstract](#)
48. Gardner AM, Martin LA, Gardner PR, *et al.*: **Steady-state and transient kinetics of *Escherichia coli* nitric-oxide dioxygenase (flavo-hemoglobin). The B₁₀ tyrosine hydroxyl is essential for dioxygen binding and catalysis.** *J Biol Chem.* 2000; **275**(17): 12581–12589.
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Igarashi J, Kobayashi K, Matsuoka A: **A hydrogen-bonding network formed by the B10–E7–E11 residues of a truncated hemoglobin from *Tetrahymena pyriformis* is critical for stability of bound oxygen and nitric oxide detoxification.** *J Biol Inorg Chem.* 2011; **16**(4): 599–609.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Ouellet Y, Millani M, Couture M, *et al.*: **Ligand interactions in the distal heme pocket of *Mycobacterium tuberculosis* truncated hemoglobin N: roles of TyrB10 and GlnE11 residues.** *Biochemistry.* 2006; **45**(29): 8770–8781.
[PubMed Abstract](#) | [Publisher Full Text](#)
51. Vinogradov SN, Hoogewijs D, Bailly X, *et al.*: **A model of globin evolution.** *Gene.* 2007; **398**(1–2): 132–142.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. Wu G, Wainwright LM, Poole RK: **Microbial globins.** *Adv Microb Physiol.* 2003; **47**: 255–310.
[PubMed Abstract](#) | [Publisher Full Text](#)
53. Pesce A, Thijs L, Nardini M, *et al.*: **HisE11 and HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of *Geobacter sulfurreducens* globin-coupled sensor.** *J Mol Biol.* 2009; **386**(1): 246–260.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Zhang W, Phillips GN Jr: **Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry.** *Structure.* 2003; **11**(9): 1097–1110.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Vasudevan SG, Armarego WL, Shaw DC, *et al.*: **Isolation and nucleotide sequence of the *hmp* gene that encodes a haemoglobin-like protein in *Escherichia coli* K-12.** *Mol Gen Genet.* 1991; **226**(1–2): 49–58.
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Gesto-Borroto R, Sánchez-Sánchez M, Arredondo-Peter R: **Dataset 1 in: A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling.** *F1000Research.* 2015.
[Data Source](#)
57. Gesto-Borroto R, Sánchez-Sánchez M, Arredondo-Peter R: **Dataset 2 in: A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling.** *F1000Research.* 2015.
[Data Source](#)
58. Gesto-Borroto R, Sánchez-Sánchez M, Arredondo-Peter R: **Dataset 3 in: A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling.** *F1000Research.* 2015.
[Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 08 July 2015

doi:10.5256/f1000research.6858.r9023



Paul Twigg

Department of Biology, University of Nebraska at Kearney, Kearney, NE, USA

I find this paper by Gesto-Borroto *et al.* to be well written and analyzed. This paper fills a gap in the knowledge base for what is known about bacterial or more specifically rhizobial globins. The authors seem to have taken great care to analyze all available globins from various rhizobial species and biovars. I have no major revisions for this work. It is comprehensive and demonstrates some interesting points about globins in rhizobia. The coexistence of various globins in the bacteria begs questions about their control and functions that undoubtedly other researchers will address. The work with the promoter analysis was particularly interesting to me indicating more than once that the globin expression is likely tied to nitrate/nitrogen metabolism. The absence of the -10 promoter area was also unexpected. The amino acid sequence alignment data also shows interesting information about the conserved positions in the sequence. The phenetic relationships are also interesting and appropriately analyzed. The structural modeling is also well done and reveals interesting points about the structure and function of the various globins. Lastly, the authors back up some of their proposals with spectroscopic data from globing extracts. Again, overall I thought that the work was well done and needs no major revisions.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 08 Jul 2015

Raul Arredondo-Peter, Universidad Autonoma del Estado de Morelos, Mexico

We thank Dr. Twigg for evaluating this article and constructive comments.

Competing Interests: No competing interests were disclosed.

Referee Report 26 May 2015

doi:10.5256/f1000research.6858.r8649



Manuel Becana

Department of Plant Nutrition, Spanish National Research Council, Zaragoza, Spain

This is a well-written paper on a subject of great interest. There is very little information of rhizobial globins and the authors have done a good job by systematically analyzing the composition of globin genes of 62 genomes in various genera, species and biovars of rhizobia. The authors are experts in the phylogeny and evolution of plant hemoglobins, and I have no major comments to improve this work. It will nevertheless be of interest for future work to address the issue of why several types of hemoglobins coexist in rhizobia. For example, both truncated hemoglobins and flavohemoglobins seem to be present within the same species and strain, although this would have to be verified by identifying the proteins themselves rather than by only gene sequencing or by analyzing differential spectra (reduced + CO vs reduced) in bacterial extracts. Both classes of hemoglobins have been proposed to act as modulators of NO concentration, but they are unlikely to have redundant functions. An interesting, additional aspect of the work is the mapping analysis, including the report of flanking sequences. This hints to a role of at least some rhizobial globins in nitrogen metabolism. This observation is very timing because of the recent discovery that truncated hemoglobins of *Chlamydomonas* regulate nitrate reductase.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 26 May 2015

Raul Arredondo-Peter, Universidad Autonoma del Estado de Morelos, Mexico

We thank Dr. Becana for evaluating this article and constructive comments.

Competing Interests: No competing interests were disclosed.
