International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 6, No. 2 (2017): 922-933
Research Article
Open Access

ISSN: 2320-0634

Prediction of folding pathway and rate for selected rhizobial single domain and truncated hemoglobins using an average distance map method

Masanari Matsuoka[1], Michiro Kabata[1], Kohei Ohnishi[1], Takeshi Kikuchi[1] and Raúl Arredondo-Peter[2]*

[1] *Department of Bioinformatics, College of Life Science, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan*
[2] *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, Avenida Universidad 1001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, México.*

*Corresponding author: Raúl Arredondo-Peter; email: ra@uaem.mx

Received: 13 February 2017 Accepted: 27 February 2017 Online: 02 March 2017

ABSTRACT

Hemoglobins (Hbs) are proteins widely distributed in organisms from the three kingdoms of life. Genomic analysis revealed that genes coding for single domain Hbs (SDgbs), flavohemoglobins, globin-coupled sensors and truncated Hbs (tHbs) exist in rhizobial bacteria. Rhizobial Hb sequences have been characterized using bioinformatics methods, however nothing is known about the folding pathway and rate of rhizobial SDgbs and tHbs. Here, we report the prediction of folding pathway and rate for selected rhizobial SDgbs and tHbs using an Average Distance Map method. Results predicted that folding of most of the rhizobial SDgbs and tHbs analyzed in this work occurs throughout the formation of two compact modules, that helix composition for compact modules is rather variable and that protein folding mostly occurs at moderate rate either in the N→C or C→N direction.

Keywords: ADM method, molecular dynamics, *Rhizobium*, unfolding, 2/2, 3/3.

Abbreviations: ADM, Average Distance Map method; Hb, hemoglobin; Mb, myoglobin; SDGb, single-domain hemoglobin; tHb, truncated hemoglobin

1. INTRODUCTION

Hemoglobins (Hbs) are proteins structurally characterized by a particular arrangement of 6 to 8 α-helices (designated with letters A to H) known as the "globin-fold". This protein folding forms a hydrophobic pocket where a heme prosthetic group is located [1]. Proximal side of heme-Fe of Hbs is coordinated by a protein´s His whereas a variety of gaseous ligands (*e.g.* O_2 and NO) reversibly bind to the distal side of heme-Fe. Two structural types of the globin fold have been identified in Hbs [2]: the 2/2- and 3/3-fold. In the 2/2-Hbs helices B and E overlap to helices G and H, and in the 3/3-Hbs helices A, E and F overlap to helices B, G and H. Also, three evolutionary families have been identified in Hbs [3, 4]: the M, S and T Hb families. The M Hbs include flavohemoglobins and single-domain Hbs (SDgbs), the S Hbs include globin-coupled sensors, protoglobins and sensor single-domain Hbs, and the T Hbs include truncated Hbs (tHbs), which are further classified into class 1, class 2 and class 3 tHbs. The M and S Hbs fold into the 3/3-fold whereas the tHbs fold into the 2/2-fold.

Hemoglobins are widely distributed in organisms from the three kingdoms of life, *i.e.* in Archaea, Eubacteria and Eukarya [2]. A comprehensive genomic analysis revealed that *hb* genes belonging to the M, S and T families exist in the genomes of 1185 Eubacteria, including several rhizobial genomes [4]. Rhizobia comprise a group of bacteria that establish symbiotic

relationships with leguminous plants where the fixation of atmospheric N_2 occurs. A bioinformatics analysis of rhizobial Hb sequences revealed that M, S and T Hb families exist in rhizobia and that predicted rhizobial SDgbs and tHbs fold into the 3/3- and 2/2-fold, respectively [5]. However, nothing is known about the folding pathway and rate of rhizobial SDgbs and tHbs.

A variety of experimental methods has been used to elucidate protein folding. These methods include UV-Vis and CD spectroscopy, fluorometry coupled to mass spectroscopy and NMR relaxation dispersion [6-12]. However, a computational method based on average contact maps within a protein sequence was developed to predict the existence and location of folding cores that could function as nucleation sites for protein folding. This method is known as the Average Distance Map (ADM) method [13, 14]. The ADM results for the prediction of the sperm whale myoglobin (Mb) and soybean leghemoglobin folding [15] were consistent with experimental data [16]. Thus, ADM is a reliable method to predict the folding pathway and rate of proteins. A detailed description of the ADM method is reported by Matsuoka *et al*. [17] (see also the Materials and Methods section). Briefly, the ADM method generates a contact map from the average distance of Cα atoms in a protein sequence and from the contacts of various loop lengths. Also, the ADM method calculates η values that measure the strength of a compact region or domain/subdomain into the tertiary structure of a protein. A high η value predicts that a stable compact region could form during protein folding. A protein formed by domains/subdomains with similar η values could fold faster than a protein formed by domains/subdomains with different η values. Thus, rates for protein folding are classified by the ADM method as fast or moderate based on whether the η values are similar or different, respectively. Here, we report the prediction of folding pathway and rate for selected rhizobial SDgbs and tHbs using the ADM method.

Table 1: Acronyms for the rhizobial SDgb and tHb sequences analyzed in this work. See the Materials and Methods section for explanation.

Species	Strain	Hb acronym
SDgbs		
Azorhizobium doebereinerae	UFLA1-100	AzodoeUFLA1-100SDgb
Bradyrhizobium elkanii	USDA3254	BraelkUSDA3254SDgb1
		BraelkUSDA3254SDgb2
	USDA3259	BraelkUSDA3259SDgb1
	USDA94	BraelkUSDA94SDgb2
	WSM1741	BraelkWSM1741SDgb2
Bradyrhizobium japonicum	USDA124	BrajapUSDA124SDgb1
	USDA38	BrajapUSDA38SDgb2
tHbs		
Class 1		
Mesorhizobium ciceri	CMG6	MescicCMG6tHb
Cupriavidus necator	N1	CupnecN1tHb1
Bradyrhizobium elkanii	USDA94	BraelkUSDA94tHb1
Class 2		
Cupriavidus necator	N1	CupnecN1tHb2
Rhizobium lupini	HPC(L)	RhilupHPC(L)tHb2
Rhizobium leguminosarum	Vc2	RhilegVc2tHb1
Bradyrhizobium japonicum	USDA38	BrajapUSDA38tHb2
Azorhizobium doebereinerae	UFLA1-100	AzodoeUFLA1-100tHb2
Burkholderia phymatum	STM815	BurphySTM815tHb2
Rhizobium etli	KIM5	RhietlKIM5tHb
Sinorhizobium fredii	HH103	SinfreHH103tHb
Class 3		
Bradyrhizobium japonicum	USDA123	Brajap123tHb1
Bradyrhizobium elkanii	USDA76	BraelkUSDA76tHb2
Mesorhizobium loti	NZP2037	MeslotNZP2037tHb2
Rhizobium lupini	HPC(L)	RhilupHPC(L)tHb1
Sinorhizobium meliloti	1021	Sinmel1021tHb2
Burkholderia phymatum	STM815	BurphySTM815tHb1
Azorhizobium doebereinerae	UFLA1-100	AzodoeUFLA1-100tHb1
Rhizobium leguminosarum	GB30	RhilegGB30tHb2
Rhizobium etli	CNPAF512	RhietlCNPAF512tHb

2. MATERIALS AND METHODS

2.1 Protein sequences and sequence alignments

Selected rhizobial SDgb and tHb sequences representative for the major evolutionary clades of rhizobial Hbs were obtained from Gesto-Borroto *et al*. [5] (Table 1). Acronyms for the rhizobial SDgbs and tHbs correspond to the first three binomial (genus and species) letters followed by the strain name, Hb type and Hb copy number [5]. Alignment of multiple protein sequences was performed using the EMBL-EBI

program (http://www.ebi.ac.uk/) from the MView server (http://www.ebi.ac.uk/Tools/msa/mview/). Identity values were obtained from pairwise sequence alignments using the ClustalX program [18].

2.2 ADM analysis of rhizobial SDgb and tHb sequences and prediction of compact areas

Prediction of folding pathway and rate for selected rhizobial SDgb and tHb sequences were calculated using the ADM method as described elsewhere [13-15, 17]. The ADM method generated a contact map for the rhizobial SDgb and tHb sequences which is represented by compact regions indicated with light red to black bars referring to the magnitude of the η values. A region with the highest η value was defined as a domain. Regions with high η values within a domain were defined as subdomains. It was predicted that two subdomains exhibiting different η values could fold first throughout the formation of the subdomain exhibiting the higher η value and subsequently throughout the formation of the subdomain exhibiting the lower η value. Also, it was predicted that two subdomains could fold simultaneously (*i.e.* without the formation of heterogeneous intermediates) if they exhibit similar η values. Thus, the ADM method predicted that a protein formed by subdomains exhibiting similar η values could fold faster than a protein formed by subdomains exhibiting different η values. Likewise, folding of a protein exhibiting a large and strong compact region was predicted to take place without the existence of intermediates, and thus was classified as a fast folding protein.

2.3 Unfolding simulations of rhizobial SDgb and tHb sequences

Twenty nanosec equilibrium simulations and 10 nanosec unfolding simulations for the selected rhizobial SDgb and tHb sequences at 298 and 398°K, respectively, were performed using Amber 12 [19-21] in an NVT ensemble. Water molecules were placed using the placevent algorithm [22] which provides highly possible snapshots for water molecules based on the distribution calculated by 3D-RISM [23]. Heatings were calculated by employing an NPT ensemble up to 348°K.

3. RESULTS AND DISCUSSION

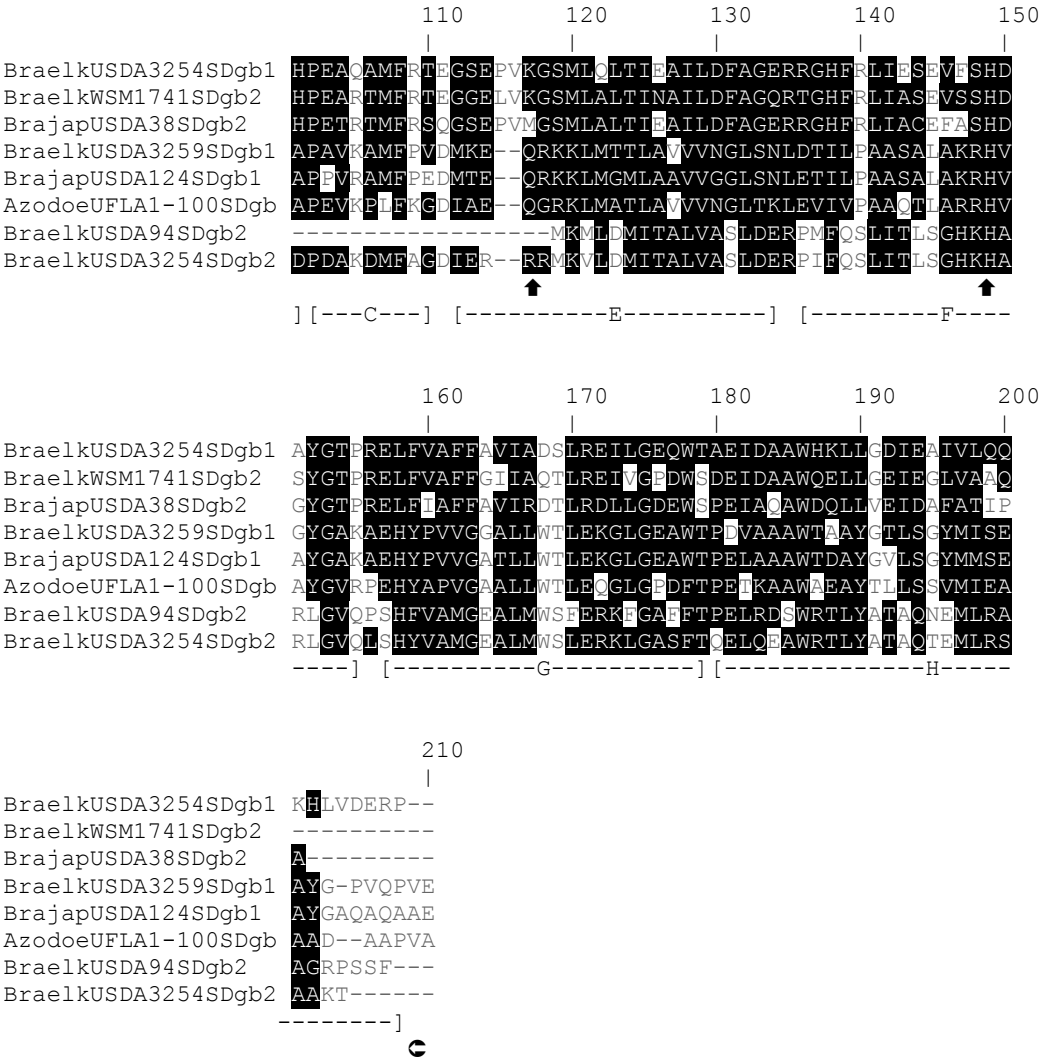
Experimental and *in silico* methods have been used to elucidate the folding of Hbs, including sperm whale Mb [16] and plant Hbs [24], respectively. The only bacterial Hb analyzed so far from the protein folding perspective is the *Escherichia coli* flavohemoglobin (Hmp) [25], which is a chimeric Hb consisting of a globin and a NAD(P)$^+$/FAD-containing reductase domains. Analysis of the apoHmp globin domain revealed that this protein folds similarly to other Hbs. Hemoglobin sequences have been detected and analyzed in a variety of rhizobia, including single domain SDgbs and tHbs and chimeric flavohemoglobins and globin-coupled sensors [4, 5]. However, nothing is known about the folding of rhizobial Hbs. Thus, we used the ADM method to predict the folding pathway and rate of 8 and 20 single domain rhizobial SDgbs and tHbs, respectively (whose 3 corresponded to tHbs class 1, 8 corresponded to tHbs class 2 and 9 corresponded to tHbs class 3) (Table 1), which are representative for the major evolutionary clades of rhizobial Hbs [5].

(A)

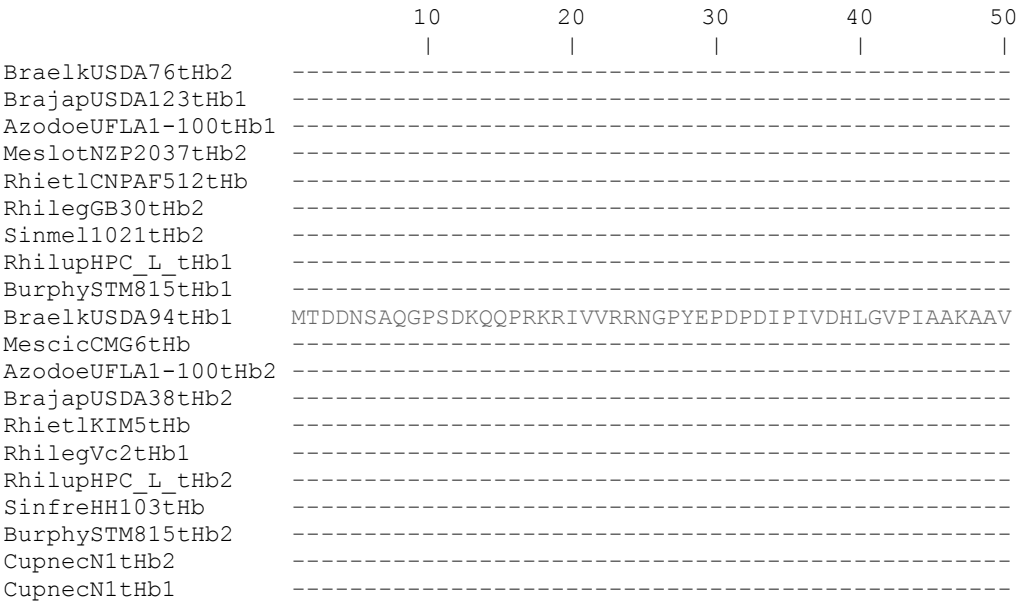
```
                                10        20        30        40        50
                                 |         |         |         |         |
BraelkUSDA3254SDgb1 ----------------------------------------------------
BraelkWSM1741SDgb2  MVYQFGLRLADRQSLVVQVSGGALTRTSSVASTIKIKSQFSRKLRVDTIF
BrajapUSDA38SDgb2   ----------------------------------------------------
BraelkUSDA3259SDgb1 ----------------------------------------------------
BrajapUSDA124SDgb1  ----------------------------------------------------
AzodoeUFLA1-100SDgb ----------------------------------------------------
BraelkUSDA94SDgb2   ----------------------------------------------------
BraelkUSDA3254SDgb2 ----------------------------------------------------

                                60        70        80        90       100
                                 |         |         |         |         |
BraelkUSDA3254SDgb1 ----------------MTPSSNPIERSFELAAAACDDLTPSVYRRLFRD
BraelkWSM1741SDgb2  EQGLVPCSFFSLRSQTTMTASSNPIERSFELAAARCEDLTPLVYRRLFEA
BrajapUSDA38SDgb2   ---------------MPMSAPLNSIENSFELAASRCADLTPLVYQRLFEQ
BraelkUSDA3259SDgb1 ----------------MNPAQIKLVQDSFGKVAPISEQAAVIFYDRLFEV
BrajapUSDA124SDgb1  ----------------MTPEQITLIQQSFAKVAPISEQAAVLFYDRLFEV
AzodoeUFLA1-100SDgb ----------------MTPSQVELVQSSFAKVAPIADTAAGLFYGRLFET
BraelkUSDA94SDgb2   ----------------------------------------------------
BraelkUSDA3254SDgb2 ----------------MTPEQVDLIGISFDAMWPIRRDIADLCYSRFVEL

                                     [-------A-------][-------B-------
                                     ➲
```


110 120 130 140 150
BraelkUSDA3254SDgb1 HPEAQAMFRTEGSEPVKGSMLQLTIEAILDFAGERRGHFRLIESEVESHD
BraelkWSM1741SDgb2 HPEARTMFRTEGGELVKGSMLALTINAILDFAGQRTGHFRLIASEVSSHD
BrajapUSDA38SDgb2 HPETRTMFRSQGSEPVMGSMLALTIEAILDFAGERRGHFRLIACEFASHD
BraelkUSDA3259SDgb1 APAVKAMFPVDMKE--QRKKLMTTLAVVVNGLSNLDTILPAASALAKRHV
BrajapUSDA124SDgb1 APPVRAMFPEDMTE--QRKKLMGMLAAVVGGLSNLETILPAASALAKRHV
AzodoeUFLA1-100SDgb APEVKPLFKGDIAE--QGRKLMATLAVVVNGLTKLEVIVPAAQTIARRHV
BraelkUSDA94SDgb2 ----------MKMLDMITALVASLDERPMFQSLITLSGHKHA
BraelkUSDA3254SDgb2 DPDAKDMFAGDIER--RRMKVLDMITALVASLDERPIFQSLITLSGHKHA
][---C---] [----------E----------] [---------F----
160 170 180 190 200
BraelkUSDA3254SDgb1 AYGTPRELFVAFFAVIADSLREILGEQWTAEIDAAWHKLLGDIEAIVLQQ
BraelkWSM1741SDgb2 SYGTPRELFVAFFGIIAQTLREIVGPDWSDEIDAAWQELLGEIEGLVAAQ
BrajapUSDA38SDgb2 GYGTPRELFIAFFAVIRDTLRDLLGDEWSPEIAQAWDQLLVEIDAFATIP
BraelkUSDA3259SDgb1 GYGAKAEHYPVVGGALLWTLEKGLGEAWTPDVAAAWTAAYGTLSGYMISE
BrajapUSDA124SDgb1 AYGAKAEHYPVVGATLLWTLEKGLGEAWTPELAAAWTDAYGVLSGYMMSE
AzodoeUFLA1-100SDgb AYGVRPEHYAPVGAALLWTLEQGLGPDFTPETKAAWAEAYTLLSSVMIEA
BraelkUSDA94SDgb2 RLGVQPSHFVAMGEALMWSFERKFGAFFTPELRDSWRTLYATAQNEMLRA
BraelkUSDA3254SDgb2 RLGVQLSHYVAMGEALMWSLERKLGASFTQELQEAWRTLYATAQTEMLRS
----] [----------G----------][--------------H-----
210
BraelkUSDA3254SDgb1 KHLVDERP--
BraelkWSM1741SDgb2 ----------
BrajapUSDA38SDgb2 A---------
BraelkUSDA3259SDgb1 AYG-PVQPVE
BrajapUSDA124SDgb1 AYGAQAQAAE
AzodoeUFLA1-100SDgb AAD--AAPVA
BraelkUSDA94SDgb2 AGRPSSF---
BraelkUSDA3254SDgb2 AAKT------
--------]

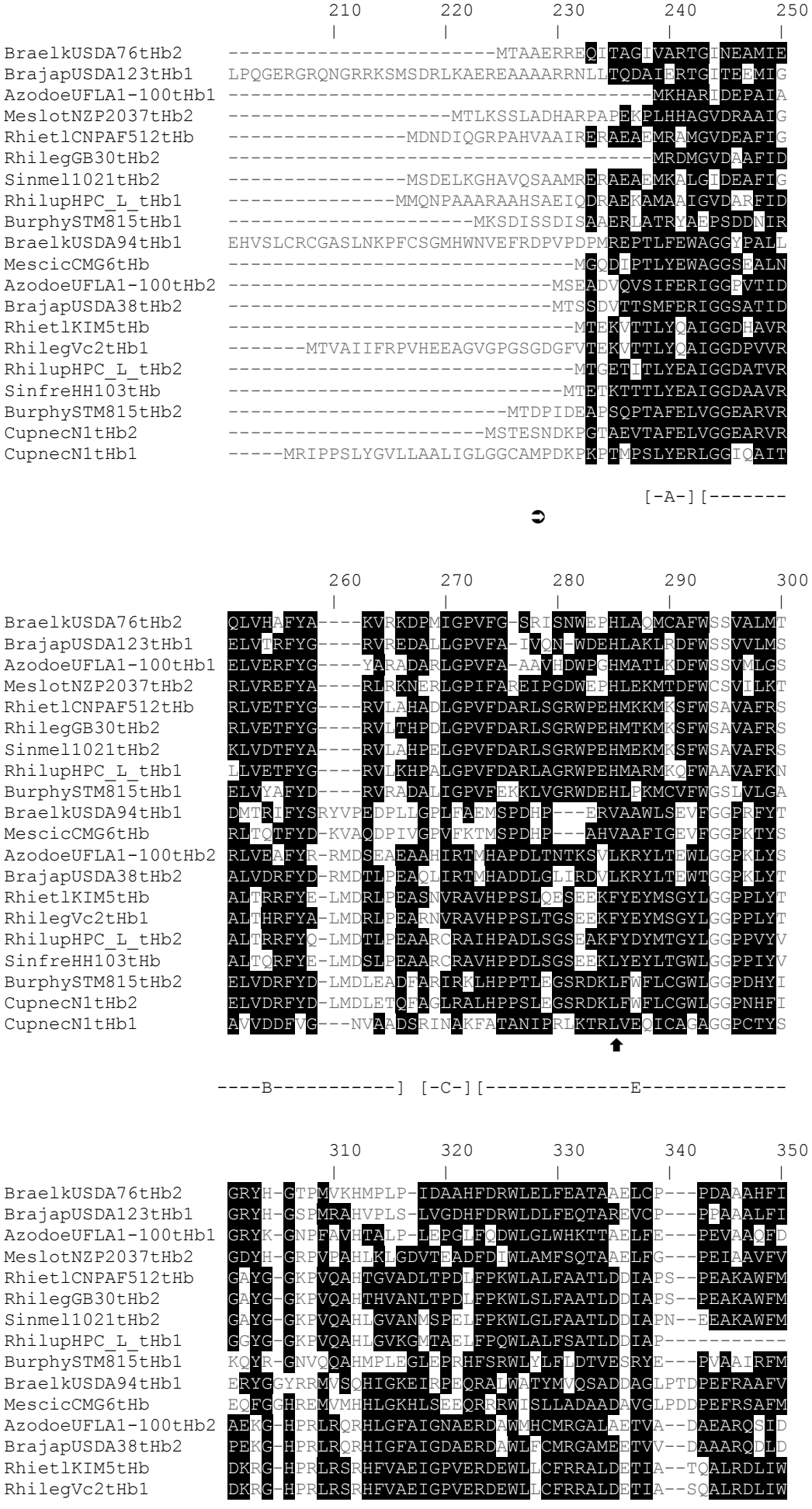
(B)

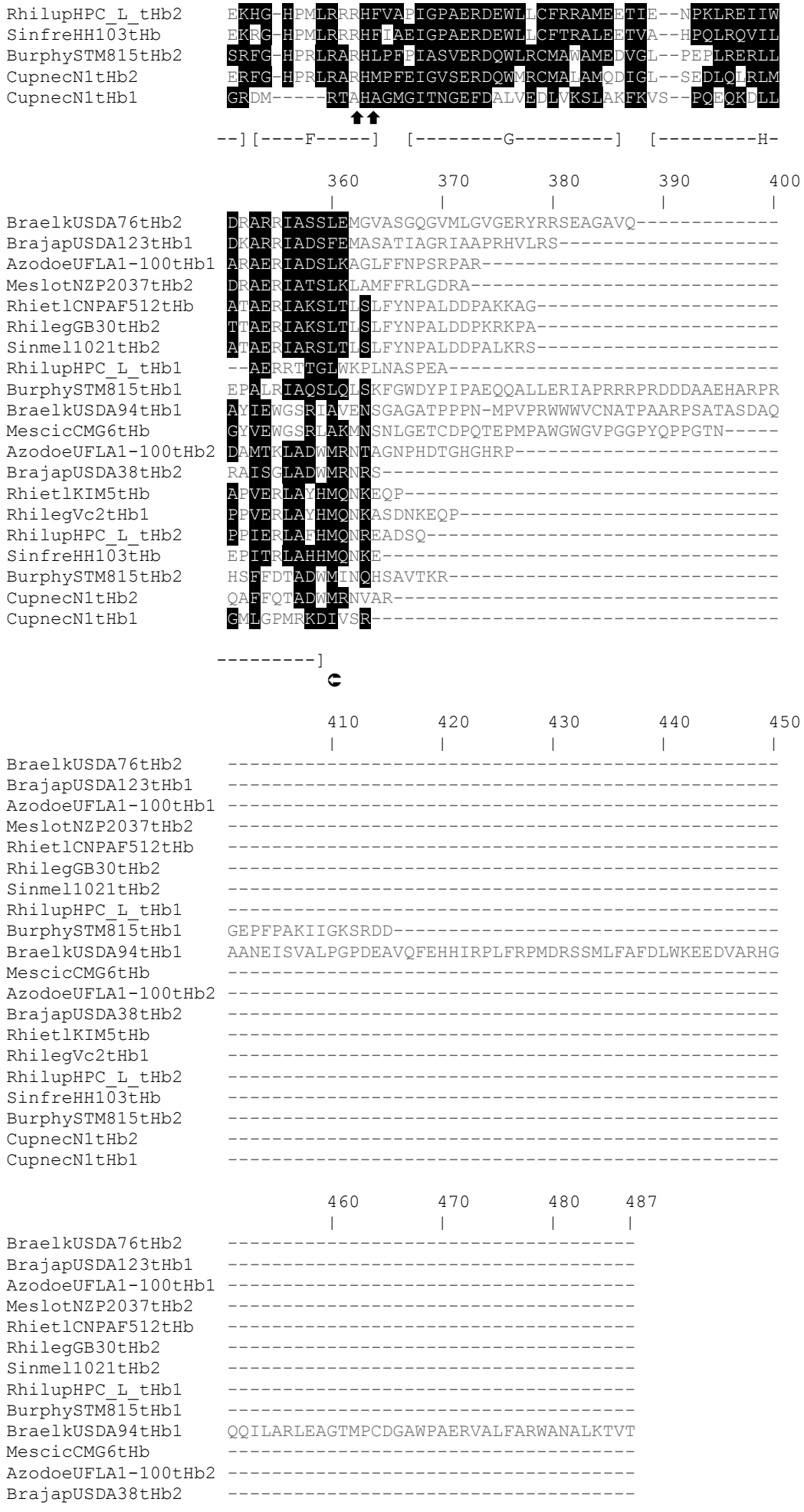

10 20 30 40 50
BraelkUSDA76tHb2 ----------
BrajapUSDA123tHb1 ----------
AzodoeUFLA1-100tHb1 ----------
MeslotNZP2037tHb2 ----------
RhietlCNPAF512tHb ----------
RhilegGB30tHb2 ----------
Sinmel1021tHb2 ----------
RhilupHPC_L_tHb1 ----------
BurphySTM815tHb1 ----------
BraelkUSDA94tHb1 MTDDNSAQGPSDKQQPRKRIVVRRNGPYEPDPDIPIVDHLGVPIAAKAAV
MescicCMG6tHb ----------
AzodoeUFLA1-100tHb2 ----------
BrajapUSDA38tHb2 ----------
RhietlKIM5tHb ----------
RhilegVc2tHb1 ----------
RhilupHPC_L_tHb2 ----------
SinfreHH103tHb ----------
BurphySTM815tHb2 ----------
CupnecN1tHb2 ----------
CupnecN1tHb1 ----------

```
                          60        70        80        90        100
                          |         |         |         |         |
BraelkUSDA76tHb2    --------------------------------------------------
BrajapUSDA123tHb1   --------------------------------------------------
AzodoeUFLA1-100tHb1 --------------------------------------------------
MeslotNZP2037tHb2   --------------------------------------------------
RhietlCNPAF512tHb   --------------------------------------------------
RhilegGB30tHb2      --------------------------------------------------
Sinmel1021tHb2      --------------------------------------------------
RhilupHPC_L_tHb1    --------------------------------------------------
BurphySTM815tHb1    --------------------------------------------------
BraelkUSDA94tHb1    RLCRCGQSQTKPFCDDSHVARGFTDARDPRRVPDRLEVYAGQQAYVFDNR
MescicCMG6tHb       --------------------------------------------------
AzodoeUFLA1-100tHb2 --------------------------------------------------
BrajapUSDA38tHb2    --------------------------------------------------
RhietlKIM5tHb       --------------------------------------------------
RhilegVc2tHb1       --------------------------------------------------
RhilupHPC_L_tHb2    --------------------------------------------------
SinfreHH103tHb      --------------------------------------------------
BurphySTM815tHb2    --------------------------------------------------
CupnecN1tHb2        --------------------------------------------------
CupnecN1tHb1        --------------------------------------------------

                          110       120       130       140       150
                          |         |         |         |         |
BraelkUSDA76tHb2    --------------------------------------------------
BrajapUSDA123tHb1   --------------------------------------------------
AzodoeUFLA1-100tHb1 --------------------------------------------------
MeslotNZP2037tHb2   --------------------------------------------------
RhietlCNPAF512tHb   --------------------------------------------------
RhilegGB30tHb2      --------------------------------------------------
Sinmel1021tHb2      --------------------------------------------------
RhilupHPC_L_tHb1    --------------------------------------------------
BurphySTM815tHb1    --------------------------------------------------
BraelkUSDA94tHb1    GTCAHSGFCTNRLASVFRLGEQPFIAPSGARLDDLINAVRRCPSGALGIG
MescicCMG6tHb       --------------------------------------------------
AzodoeUFLA1-100tHb2 --------------------------------------------------
BrajapUSDA38tHb2    --------------------------------------------------
RhietlKIM5tHb       --------------------------------------------------
RhilegVc2tHb1       --------------------------------------------------
RhilupHPC_L_tHb2    --------------------------------------------------
SinfreHH103tHb      --------------------------------------------------
BurphySTM815tHb2    --------------------------------------------------
CupnecN1tHb2        --------------------------------------------------
CupnecN1tHb1        --------------------------------------------------

                          160       170       180       190       200
                          |         |         |         |         |
BraelkUSDA76tHb2    --------------------------------------------------
BrajapUSDA123tHb1   ----------------------------MATRLTANTVAHPSRRFAASHP
AzodoeUFLA1-100tHb1 --------------------------------------------------
MeslotNZP2037tHb2   --------------------------------------------------
RhietlCNPAF512tHb   --------------------------------------------------
RhilegGB30tHb2      --------------------------------------------------
Sinmel1021tHb2      --------------------------------------------------
RhilupHPC_L_tHb1    --------------------------------------------------
BurphySTM815tHb1    --------------------------------------------------
BraelkUSDA94tHb1    IGPERNANLSDVNRPPQIEVSKDGPYRVTGHVELVDEDGASITRNAGASQ
MescicCMG6tHb       --------------------------------------------------
AzodoeUFLA1-100tHb2 --------------------------------------------------
BrajapUSDA38tHb2    --------------------------------------------------
RhietlKIM5tHb       --------------------------------------------------
RhilegVc2tHb1       --------------------------------------------------
RhilupHPC_L_tHb2    --------------------------------------------------
SinfreHH103tHb      --------------------------------------------------
BurphySTM815tHb2    --------------------------------------------------
CupnecN1tHb2        --------------------------------------------------
CupnecN1tHb1        --------------------------------------------------
```


210 220 230 240 250
BraelkUSDA76tHb2 ----------MTAAERREQITAGIVARTGINEAMIE
BrajapUSDA123tHb1 LPQGERGRQNGRRKSMSDRLKAEREAAAARRNLLTQDAIERTGITEEMIG
AzodoeUFLA1-100tHb1 ----------MKHARIDEPAIA
MeslotNZP2037tHb2 ----------MTLKSSLADHARPAPEKPLHHAGVDRAAIG
RhietlCNPAF512tHb ----------MDNDIQGRPAHVAAIRERAEAEMRAMGVDEAFIG
RhilegGB30tHb2 ----------MRDMGVDAAFID
Sinmel1021tHb2 ----------MSDELKGHAVQSAAMRERAEAEMKALGIDEAFIG
RhilupHPC_L_tHb1 ----------MMQNPAAARAAHSAEIQDRAEKAMAAIGVDARFID
BurphySTM815tHb1 ----------MKSDISSDISAAERLATRYAEPSDDNIR
BraelkUSDA94tHb1 EHVSLCRCGASLNKPFCSGMHWNVEFRDPVPDPMREPTLFEWAGGVPALL
MescicCMG6tHb ----------MGQDIPTLYEWAGGSEALN
AzodoeUFLA1-100tHb2 ----------MSEADVQVSIFERIGGPVTID
BrajapUSDA38tHb2 ----------MTSSDVTTSMFERIGGSATID
RhietlKIM5tHb ----------MTEKVTTLYQAIGGDEAVR
RhilegVc2tHb1 -------MTVAIIFRPVHEEAGVGPGSGDGFVTEKVTTLYQAIGGDPVVR
RhilupHPC_L_tHb2 ----------MTGETITLYEAIGGDATVR
SinfreHH103tHb ----------MTETKTTLYEAIGGDAAVR
BurphySTM815tHb2 ----------MTDPIDEAPSQPTAFELVGGEARVR
CupnecN1tHb2 ----------MSTESNDKPGTAEVTAFELVGGEARVR
CupnecN1tHb1 -----MRIPPSLYGVLLAALIGLGGCAMPDKPKPTMPSLYERLGGIQAIT
[-A-][-------
260 270 280 290 300
BraelkUSDA76tHb2 QLVHAFYA----KVRKDPMIGPVFG-SRISNWEPHLAQMCAFWSSVALMT
BrajapUSDA123tHb1 ELVTRFYG----RVREDALLGPVFA-IVQN-WDEHLAKLRDFWSSVVLMS
AzodoeUFLA1-100tHb1 ELVERFYG----YARADARLGPVFA-AAVHDWPGHMATLKDFWSSVMLGS
MeslotNZP2037tHb2 RLVREFYA----RLRKNERLGPIFAREIPGDWEPHLEKMTDFWCSVLLKT
RhietlCNPAF512tHb RLVETFYG----RVLAHADLGPVFDARLSGRWPEHMKKMKSFWSAVAFRS
RhilegGB30tHb2 RLVETFYG----RVLTHPDLGPVFDARLSGRWPEHMTKMKSFWSAVAFRS
Sinmel1021tHb2 KLVDTFYA----RVLAHPELGPVFDARLSGRWPEHMEKMKSFWSAVAFRS
RhilupHPC_L_tHb1 LLVETFYG----RVLKHPALGPVFDARLAGRWPEHMARMKQFWAAVAFKN
BurphySTM815tHb1 ELVYAFYD----RVRADALIGPVFEKKLVGRWDEHLPKMCVFWGSIVLGA
BraelkUSDA94tHb1 DMTRIFYSRYVPEDPLLGPLFAEMSPDHP---ERVAAWLSEVFGGPRFYT
MescicCMG6tHb RLTQTFYD-KVAQDPIVGPVFKTMSPDHP---AHVAAFIGEVFGGPKTYS
AzodoeUFLA1-100tHb2 RLVEAFYR-RMDSEAEAAHIRTMHAPDLTNTKSVLKRYLTEWLGGPKLYS
BrajapUSDA38tHb2 ALVDRFYD-RMDTLPEAQLIRTMHADDLGLIRDVLKRYLTEWTGGPKLYT
RhietlKIM5tHb ALTRRFYE-LMDRLPEASNVRAVHPPSLQESEEKFYEYMSGYLGGPPLYT
RhilegVc2tHb1 ALTHRFYA-LMDRLPEARNVRAVHPPSLTGSEEKFYEYMSGYLGGPPLYT
RhilupHPC_L_tHb2 ALTRRFYQ-LMDTLPEAARCRAIHPADLSGSEAKFYDYMTGYLGGPPVYV
SinfreHH103tHb ALTQRFYE-LMDSLPEAARCRAVHPPDLSGSEEKLYEYLTGWLGGPPIYV
BurphySTM815tHb2 ELVDRFYD-LMDLEADFARIRKLHPPTLEGSRDKLFWFLCGWLGGPDHYI
CupnecN1tHb2 ELVDRFYD-LMDLETQFAGLRALHPPSLEGSRDKLFWFLCGWLGGPNHFI
CupnecN1tHb1 AVVDDFVG---NVAADSRINAKFATANIPRLKTRLVEQICAGAGGPCTYS
----B----------] [-C-][-------------E-------------
310 320 330 340 350
BraelkUSDA76tHb2 GRYH-GTPMVKHMPLP-IDAAHFDRWLELFEATAAELCP---PDAAAHFI
BrajapUSDA123tHb1 GRYH-GSPMRAHVPLS-LVGDHFDRWLDLFEQTAREVCP---PFAAALFI
AzodoeUFLA1-100tHb1 GRYK-GNPFAVHTALP-LEPGLFQDWLGLWHKTTAELFE---PEVAAQFD
MeslotNZP2037tHb2 GDYH-GRPVPAHLKLGDVTEADFDIWLAMFSQTAAELFG---PEIAAVFV
RhietlCNPAF512tHb GAYG-GKPVQAHTGVADLTPDLFPKWLALFAATLDDIAPS--PEAKAWFM
RhilegGB30tHb2 GAYG-GKPVQAHTHVANLTPDLFPKWLSLFAATLDDIAPS--PEAKAWFM
Sinmel1021tHb2 GAYG-GKPVQAHIGVANMSPELFPKWLGLFAATLDDIAPN--PEAKAWFM
RhilupHPC_L_tHb1 GGYG-GKPVQAHLGVKGMTAELFPQWLALFSATLDDIAP----------
BurphySTM815tHb1 KQYR-GNVQQAHMPLEGLEPRHFSRWLYLFLDTVESRYE---PVAAIRFM
BraelkUSDA94tHb1 ERYGGYRRMVSQHIGKEIRPEQRALWATYMVQSADDAGLPTDPEFRAAFV
MescicCMG6tHb EQFGGHREMVMHHLGKHLSEEQRRRWISLLADAADAVGLPDDPEFRSAFM
AzodoeUFLA1-100tHb2 AEKG-HPRLRQRHLGFAIGNAERDAWMHCMRGALAETVA--DAEARQSID
BrajapUSDA38tHb2 PEKG-HPRLRQRHIGFAIGDAERDAWLFCMRGAMEETVV--DAAARQDLD
RhietlKIM5tHb DKRG-HPRLRSRHFVAEIGPVERDEWLLCFRRALDETIA--TQALRDLIW
RhilegVc2tHb1 DKRG-HPRLRSRHFVAEIGPVERDEWLLCFRRALDETIA--SQALRDLIW


RhilupHPC_L_tHb2
SinfreHH103tHb
BurphySTM815tHb2
CupnecN1tHb2
CupnecN1tHb1
--][----F-----] [--------G---------] [---------H-
360 370 380 390 400
BraelkUSDA76tHb2
BrajapUSDA123tHb1
AzodoeUFLA1-100tHb1
MeslotNZP2037tHb2
RhietlCNPAF512tHb
RhilegGB30tHb2
Sinmel1021tHb2
RhilupHPC_L_tHb1
BurphySTM815tHb1
BraelkUSDA94tHb1
MescicCMG6tHb
AzodoeUFLA1-100tHb2
BrajapUSDA38tHb2
RhietlKIM5tHb
RhilegVc2tHb1
RhilupHPC_L_tHb2
SinfreHH103tHb
BurphySTM815tHb2
CupnecN1tHb2
CupnecN1tHb1
---------]
410 420 430 440 450
460 470 480 487

```
RhietlKIM5tHb       ------------------------------------
RhilegVc2tHb1       ------------------------------------
RhilupHPC_L_tHb2    ------------------------------------
SinfreHH103tHb      ------------------------------------
BurphySTM815tHb2    ------------------------------------
CupnecN1tHb2        ------------------------------------
CupnecN1tHb1        ------------------------------------
```

Figure 1. Sequence alignment of selected rhizobial SDgbs (A) and tHbs (B). Homologous regions are shown with black background. Vertical arrows indicate distal amino acids and proximal His at SDgbs and tHbs positions 117 and 149 and 285 and 312/313, respectively. Limits for globin domains are indicated with right- and left-oriented arrows within black circles. Helices are indicated with letters A to H in the SDgb and tHb sequences based on the *Vitreoscilla* SDgb [26] and *Mycobacterium tuberculosarum* tHb [27] structures, respectively.

Table 2: Sequence identity (%) between the rhizobial SDgbs (A) and tHbs (B) analyzed in this work. Values were calculated by pairwise sequence alignments from sequences aligned in figure 1. Lowest and highest values are indicated with gray background.

(A)

	AzodoeUFLA1-100SDgb	BraelkUSDA94SDgb2	BraelkUSDA3254SDgb1	BraelkUSDA3254SDgb2	BraelkUSDA3259SDgb1	BraelkWSM1741SDgb2	BrajapUSDA38SDgb2	BrajapUSDA124SDgb1
AzodoeUFLA1-100SDgb								
BraelkUSDA94SDgb2	16.1							
BraelkUSDA3254SDgb1	20.8	12.7						
BraelkUSDA3254SDgb2	32.6	50.3	21.1					
BraelkUSDA3259SDgb1	58.1	13.3	20.1	28.1				
BraelkWSM1741SDgb2	14.4	6.7	48.0	13.2	14.3			
BrajapUSDA38SDgb2	20.9	11.2	60.8	20.1	20.8	43.7		
BrajapUSDA124SDgb1	57.0	14.7	22.2	31.6	79.5	15.2	22.7	

(B)

	CupnecN1tHb1	AzodoeUFLA1-100tHb2	BraelkUSDA94tHb1	BrajapUSDA38tHb2	BurphySTM815tHb2	CupnecN1tHb2	MescicCMG6tHb	RhietlKIM5tHb	RhilegVc2tHb1	RhilupHPC(L)tHb2	SinfreHH103tHb	AzodoeUFLA1-100tHb1	BraelkUSDA76tHb2	BrajapUSDA123tHb1	BurphySTM815tHb1	MeslotNZP2037tHb2	RhietlCNPAF512tHb	RhilegGB30tHb2	RhilupHPC(L)tHb1	Sinmel1021tHb2
CupnecN1tHb1																				
AzodoeUFLA1-100tHb2	15.8																			
BraelkUSDA94tHb1	4.9	5.5																		
BrajapUSDA38tHb2	16.4	62.2	6.3																	
BurphySTM815tHb2	15.4	34.2	5.5	3.9																
CupnecN1tHb2	12.5	34.0	5.9	37.6	67.8															
MescicCMG6tHb	13.3	16.9	16.3	14.5	17.7	15.2														
RhietlKIM5tHb	11.2	34.9	6.3	39.8	33.8	33.5	16.3													
RhilegVc2tHb1	12.5	29.0	5.9	32.0	30.1	29.1	13.2	75.6												
RhilupHPC(L)tHb2	11.1	34.9	5.3	44.4	37.5	34.5	16.9	66.9	54.9											
SinfreHH103tHb	14.5	41.2	5.7	46.5	39.7	38.4	16.3	71.2	57.1	73.1										
AzodoeUFLA1-100tHb1	8.5	12.6	4.9	11.9	10.9	9.4	13.6	12.1	9.7	11.4	11.3									
BraelkUSDA76tHb2	9.0	15.0	5.1	13.7	10.6	9.2	14.9	11.8	10.8	11.8	12.5	32.6								
BrajapUSDA123tHb1	8.1	12.0	6.5	12.0	12.0	10.9	12.2	9.0	9.0	9.0	9.5	30.2	38.6							
BurphySTM815tHb1	3.9	8.4	4.5	10.0	11.5	11.4	12.6	9.4	8.9	10.5	10.5	24.8	25.9	24.0						
MeslotNZP2037tHb2	9.8	13.9	5.3	13.4	9.3	9.9	12.2	13.4	11.8	14.0	14.0	37.9	36.4	27.8	24.0					
RhietlCNPAF512tHb	9.5	12.9	5.1	9.4	10.6	11.8	14.2	13.8	13.1	15.7	13.8	36.7	27.4	27.6	26.5	34.1				
RhilegGB30tHb2	8.3	13.3	4.7	10.8	9.9	11.1	16.7	13.7	12.4	15.8	13.6	39.8	28.8	24.6	25.2	31.7	76.7			
RhilupHPC(L)tHb1	7.5	8.6	4.9	9.2	9.9	9.8	11.9	12.5	11.9	12.3	10.5	27.3	24.3	22.1	19.7	30.2	55.7	49.3		
Sinmel1021tHb2	7.7	10.4	5.5	10.6	11.3	12.4	15.4	14.4	14.2	14.4	13.2	35.4	29.8	24.6	24.4	32.9	79.3	69.0	53.2	

3.1 General folding pathway and rate for rhizobial Hbs

Sequence identity among rhizobial SDgbs and tHbs analyzed in this work varies from 3.9 to 79.5% (Table 2), however their similarity (Fig. 1) and predicted tertiary structures [5] are highly conserved. Our analysis of the folding process of rhizobial SDgbs and tHbs using the ADM method predicted that 1 to 3 domains and 0 to 3 subdomains could exist in these proteins. Also, the domain/subdomain η values for the rhizobial SDgbs and tHbs analyzed in this work ranged from 0.053 to 0.509 for CupnecN1tHb2 and BraelkUSDA94SDgb2, respectively (Table 3). Models predicted that domain comprising residues 109 to 124 of CupnecN1tHb2 and domain comprising residues 1 to 82 of BraelkUSDA94SDgb2 were, respectively, the least and most stable compact folding modules for the rhizobial SDgbs and tHbs analyzed in this work. The ADM analysis also predicted that folding of most of the rhizobial SDgbs and tHbs occurs throughout the

formation of two compact modules, and that folding of BraelkUSDA94SDgb2, MescicCMG6tHb, BraelkUSDA94tHb1, BurphySTM815tHb2 and RhietlCNPAF512tHb occurs throughout the formation of a single compact module and that folding of RhietlKIM5tHb and MeslotNZP2037tHb2 occurs throughout the formation of three compact modules. Helix composition for the predicted rhizobial SDgb and tHb compact modules was rather variable (Table 3). Moreover, unfolding simulations showed that most of the helices from the rhizobial SDgbs and tHbs analyzed in this work maintain their structures within 20 nanosec and that helix to helix interaction during protein folding is variable (not shown).

Protein folding pathway and rate for rhizobial SDgbs and tHbs were predicted by identifying compact regions based on η values (Table 3) and ADM contact maps (Fig. 2 and 3). Results predicted that the rhizobial SDgbs and tHbs analyzed in this work fold either in the N→C or C→N directions. Also, qualitative analysis of subdomains stability predicted that, with the exception of AzodoeUFLA1-100SDgb which folds at fast rate, the rhizobial SDgbs and tHbs analyzed in this work fold at moderate rate (Table 3).

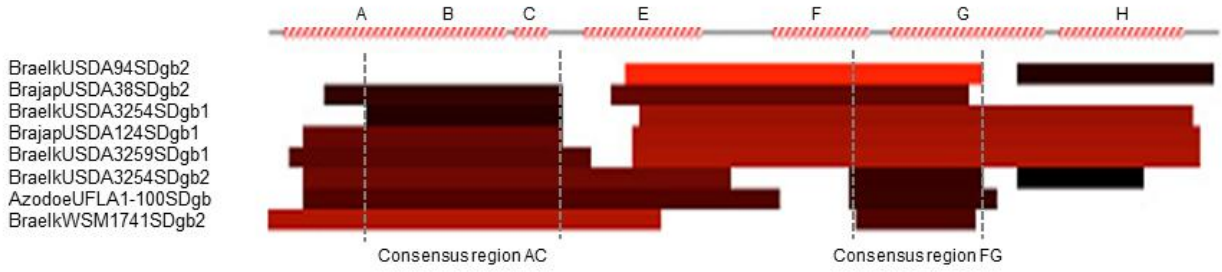


Figure 2. Location of predicted compact regions for selected rhizobial SDgbs using the ADM method. Compact regions are indicated with light red to black bars referring to the magnitude of its η value (*i.e.* light red and black correspond to high and low η values, respectively) (Table 3). Curled line above bars denotes the inferred location of α-helices by the ADM method. Vertical dashed lines indicate approximate limits for consensus regions.

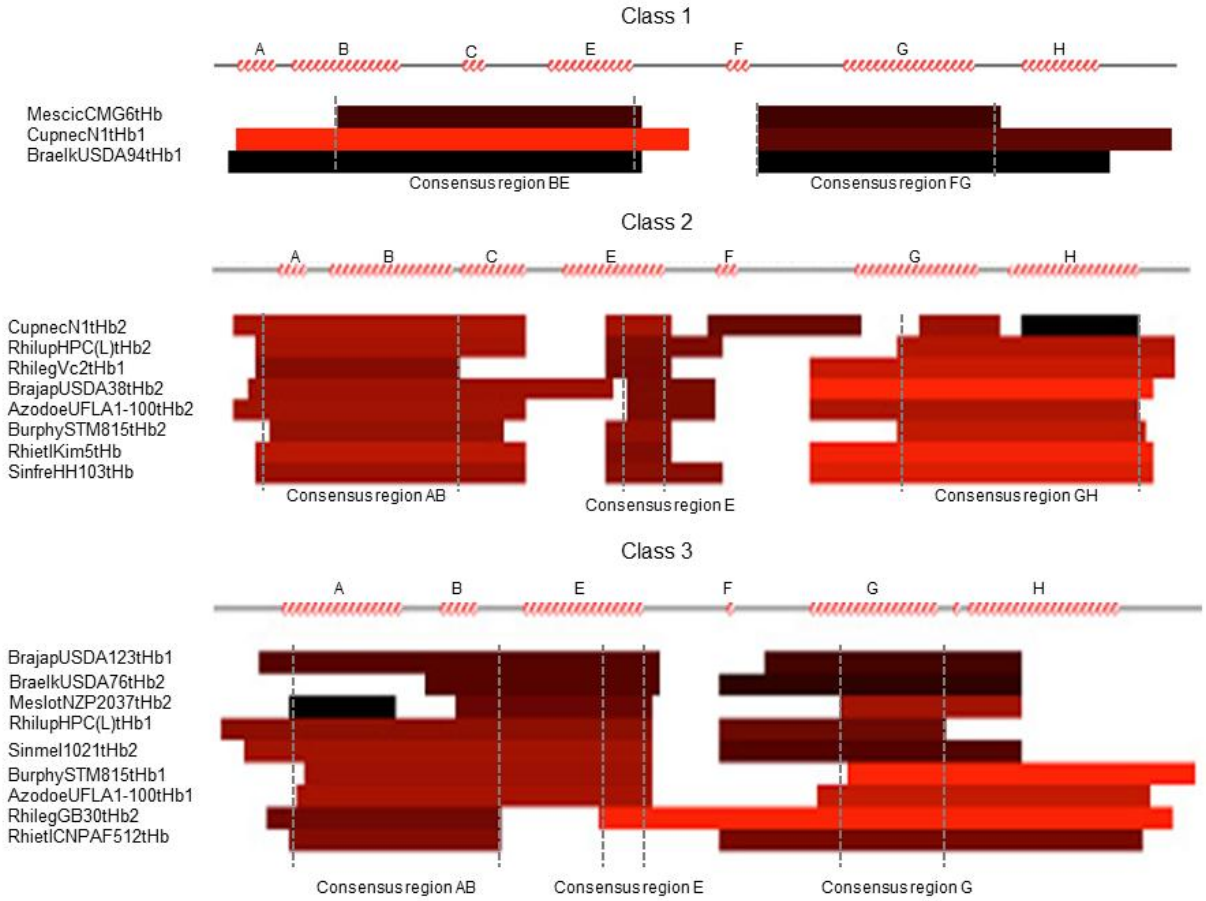


Figure 3. Location of predicted compact regions for selected rhizobial tHbs class 1, class 2 and class 3 using the ADM method. Compact regions are indicated with light red to black bars referring to the magnitude of its η value (*i.e.* light red and black correspond to high and low η values, respectively) (Table 3). Curled line above bars denotes the inferred location of α-helices by the ADM method. Vertical dashed lines indicate approximate limits for consensus regions.

Table 3: Predicted domain and subdomain location and folding pathways and rates for selected rhizobial SDgbs and tHbs.

Hb	Domain(s) (η value)[1]	Subdomains (η value)[1]	Folding pathway and rate[2]
SDgbs			
BraelkUSDA94SDgb2	1-82(0.509)	1-49(0.5), 55-82(0.178)	U→EFG→N
BrajapUSDA38SDgb2	45-95(0.272), 4-37(0.2)		U→EFG→ABC→N
BraelkUSDA3254SDgb1	46-124(0.353), 7-34(0.177)		U→EFGH→BC→N
BrajapUSDA124SDgb1	6-131(0.417)	53-131(0.368), 6-42(0.275)	U→EFGH→ABC→N
BraelkUSDA3259SDgb1	4-131(0.427)	53-131(0.377), 4-46(0.257)	U→EFGH→ABC→N
BraelkUSDA3254SDgb2	6-64(0.286), 82-100(0.208), 106-123(0.120)		U→ABCE→FG/H→N
AzodoeUFLA1-100SDgb	6-102(0.246)	6-71(0.243), 82-102(0.205)	U→[ABCE→FG]→N
BraelkWSM1741SDgb2	1-56(0.384), 85-101(0.238)		U→ABCE→FG→N
tHbs			
Class 1			
MescicCMG6tHb	15-52(0.216), 68-98(0.215)		U→BE/G→N
CupnecN1tHb1	4-116(0.333)	4-61(0.302), 66-116(0.229)	U→AB(C)E→GH→N
BraelkUSDA94tHb1	1-53(0.184), 69-113(0.184)		U→AB(C)E/GH→N
Class 2			
CupnecN1tHb2	1-40(0.289), 52-60(0.279), 95-105(0.267), 66-86(0.200), 109-124(0.053)		U→ABCE/G→H→N
RhilupHPC(L)tHb2	5-130(0.302)	93-130(0.300), 5-41(0.281), 53-68(0.227)	U→ABC/GH→E→N
RhilegVc2tHb1	82-131(0.326), 6-33(0.235), 54-62(0.221)		U→GH→AB/E→N
BrajapUSDA38tHb2	83-129(0.401), 6-55(0.270), 58-69(0.224)		U→GH→ABCE→N
AzodoeUFLA1-100tHb2	83-127(0.293), 4-43(0.275), 58-69(0.224)		U→ABC→E/GH→N
BurphySTM815tHb2	95-128(0.318), 9-40(0.258), 55-63(0.256)		U→ABCE/GH→N
RhietlKIM5tHb	81-127(0.385), 5-41(0.307), 53-61(0.234)		U→GH→ABC→E→N
SinfreHH103tHb	82-128(0.355), 6-42(0.268), 54-69(0.244)		U→GH→ABC/E→N
Class 3			
BrajapUSDA123tHb1	6-56(0.189), 71-102(0.170)		U→ABCE→G→N
BraelkUSDA76tHb2	28-57(0.189), 66-103(0.149)		U→CE→FG→N
MeslotNZP2037tHb2	82-104(0.263), 31-56(0.206), 9-22(0.103)		U→G→CE→AB→N
RhilupHPC(L)tHb1	1-57(0.242), 67-96(0.210)		U→ABCE→FG→N
Sinmel1021tHb2	3-56(0.263), 66-105(0.187)		U→ABCE→FG→N
BurphySTM815tHb1	85-129(0.352), 13-58(0.259)		U→GH→BCE→N
AzodoeUFLA1-100tHb1	8-117(0.299)	75-117(0.298), 8-53(0.267)	U→GH→BCE→N
RhilegGB30tHb2	47-122(0.349), 3-33(0.215)		U→EFGH→ABC→N
RhietlCNPAF512tHb	7-34(0.230), 64-119(0.222)		U→ABC/FGH→N

[1]Numbering corresponds to amino acids position within the globin domain (Fig. 1); η value denotes the strength of the domain formation [15]. [2]U and N correspond to unfolded and native conformations, respectively; square brackets indicate that folding might occur within short time (*e.g.* at a fast rate).

3.2 Predicted folding pathway and rate for rhizobial SDgbs

Size of most of the rhizobial SDgbs analyzed in this work is ~140 amino acids. However, BraelkWSM1741SDgb2 is longer than other rhizobial SDgbs because of the existence of extra 67 amino acids at the N-terminal. Also, BraelkUSDA94SDgb2 is shorter than other rhizobial SDgbs because of the existence of a 52 amino acids-deletion comprising helices A, B and C and part of helix E (Fig. 1A). Sequence identity among the rhizobial SDgbs analyzed in this work varies from 6.7 to 79.5% (Table 2A). Apparently, the lowest (6.7%) identity detected among rhizobial SDgbs resulted from a combined effect of the existence of extra amino acids and amino acids-deletion in BraelkWSM1741SDgb2 and BraelkUSDA94SDgb2, respectively. However,

similarity (Fig. 1A) and predicted 3/3-fold [5] for the rhizobial SDgbs analyzed in this work are highly conserved.

Table 3 shows that 1 to 3 domains could exist in the rhizobial SDgbs analyzed in this work, that η values ranged from 0.12 to 0.509 for BraelkUSDA3254SDgb2 and BraelkUSDA94SDgb2, respectively, that 2 subdomains could exist in BraelkUSDA94SDgb2, BrajapUSDA124SDgb1, BraelkUSDA3259SDgb1 and AzodoeUFLA1-100SDgb and that apparently subdomains do not exist in BrajapUSDA38SDgb2, BraelkUSDA3254SDgb1, BraelkUSDA3254SDgb2 and BraelkWSM1741SDgb2. The ADM method also predicted the existence of consensus folding regions for the rhizobial SDgbs analyzed in this work which comprise helices A, B and C and helices F and G (consensus regions AC and FG, respectively) (Fig. 2). With the exception of BraelkWSM1741SDgb2, the most stable regions of rhizobial SDgbs comprise helices E to G/H (red bars in figure 2).

The above results suggest that folding of BraelkUSDA94SDgb2, BrajapUSDA38SDgb2, BraelkUSDA3254SDgb1, BrajapUSDA124SDgb1 and BrajapUSDA3259SDgb1 could occur at moderate rate throughout the formation of a stable module comprising helices E, F and G/H followed by the formation of a module comprising helices A and B (*i.e*. these proteins could fold in the C→N direction). In contrast, folding of BraelkUSDA3254SDgb2, AzodoeUFLA1-100SDgb and BraelkWSM1741SDgb2 could occur either at fast or moderate rate throughout the formation of a module comprising helices A, B and E followed by the formation of a module comprising either helix G or helices G and H (*i.e*. these proteins could fold in the N→C direction).

3.3 Predicted folding pathway and rate for rhizobial tHbs

Size of most of the rhizobial tHbs analyzed in this work is ~130 amino acids. However, BrajapUSDA123tHb1 and BraelkUSDA94tHb1 are longer than other rhizobial tHbs because of the existence of extra ~50 amino acids at the N-terminal and ~230 and ~90 amino acids at the N- and C-terminal, respectively (Fig. 1B). Sequence identity among the rhizobial tHbs analyzed in this work varies from 3.9 to 79.3% (Table 2B). The lowest (3.9%) identity detected among rhizobial tHbs corresponded to the BurphySTM815tHb2-BrajapUSDA38tHb2 and BurphySTM815tHb1-CupnecN1tHb1 pairs but not to the BrajapUSDA123tHb1 and BraelkUSDA94tHb1 sequences. Thus, the lowest identity detected among the rhizobial tHbs analyzed in this work did not result from the existence of extra amino acids in BrajapUSDA123tHb1 and BraelkUSDA94tHb1 but from an intrinsic variability of rhizobial tHbs. Nonetheless, similarity (Fig. 1B) and predicted 2/2-fold [5] for the rhizobial tHbs analyzed in this work are highly conserved.

Table 3 shows that 1 to 3 domains could exist in the rhizobial tHbs analyzed in this work, that η values ranged from 0.053 to 0.401 for CupnecN1tHb2 and BrajapUSDA38tHb2, respectively, that apparently subdomains do not exist in most of the rhizobial tHbs analyzed in this work and that 2 and 3 subdomains could exist in CupnecN1tHb1 and AzodoeUFLA1-100tHb1 and RhilupHPC(L)tHb2, respectively. The ADM method also predicted the existence of consensus folding regions for the rhizobial tHbs analyzed in this work which comprise helices B, C and E and F and G in rhizobial tHbs class 1, helices A and B, E and G and H in rhizobial tHbs class 2 and helices A and B, E and G in rhizobial tHbs class 3 (Fig. 3). Predicted stability for the consensus folding regions of rhizobial tHbs was: tHbs class 2 > tHbs class 3 > tHbs class 1 (light red to black bars in figure 3).

The ADM results reported in table 3 and figure 3 predicted the following folding processes for the rhizobial tHbs analyzed in this work. For rhizobial tHbs class 1, folding of MescicMG6tHb and BraelkUSDA94tHb1 could occur throughout the formation of low stability single modules comprising helices B, C, E and G and helices A, B, C, E, G and H, respectively. In contrast, folding of CupnecN1tHb1 could occur throughout the formation of a stable module comprising helices A, B, C and E followed by the formation of a module comprising helices G and H. Thus, the rhizobial tHbs class 1 analyzed in this work could fold at moderate rate in the N→C direction. For rhizobial tHbs class 2, folding of BurphySTM815tHb2 could occur at moderate rate in the N→C direction throughout the formation of a stable single module comprising helices A to H. Folding of other rhizobial tHbs class 2 could occur at moderate rate either in the N→C or C→N direction throughout the formation of stable modules comprising either helices A, B and E or G and H. For rhizobial tHbs class 3, protein folding could occur at moderate rate either in the N→C or C→N direction throughout the formation of modules comprising either helices A, B and E or F, G and H. However, folding of RhietlCNPAF512tHb could occur throughout the formation of a single module comprising helices A to H and folding of MeslotNZP2037tHb2 could occur throughout the formation of modules comprising helices G, E and A and B.

4. CONCLUSION

Results from this work predicted that folding of most of the rhizobial SDgbs and tHbs and other Hbs is similar, *i.e*. Hb folding mostly occurs throughout the formation of two compact modules [6, 7, 9, 10, 12, 16, 24, 25]. However, helix composition for predicted compact modules of the rhizobial SDgbs and tHbs analyzed in this work is rather variable (Table 3). Also, our results predicted that protein folding for the rhizobial SDgbs and tHbs analyzed in this work mostly occurs at moderate rate either in the N→C or C→N direction. However, in spite of predicted variations in helix composition for compact modules and direction of

protein folding the predicted tertiary structure for rhizobial SDgbs and tHbs [5] still corresponds to the 3/3- and 2/2-fold, respectively.

Acknowledgements: 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, to RA-P.

5. REFERENCES

1. Dickerson, R. E. and Geis, I. (1983) Hemoglobin: structure, function, evolution, and pathology., The Benjamin/Cummings Pub. Co., Inc., Menlo Park, California.
2. Vinogradov, S. N., Hoogewijs, D., Bailly, X., et al. (2005) Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life., Proc. Natl. Acad. Sci. USA. 102, 11385-11389.
3. Vinogradov, S. N., Hoogewijs, D., Bailly, X., et al. (2006) A phylogenomic profile of globins., BMC Evol. Biol. 6, 31-47.
4. Vinogradov, S. N., Tinajero-Trejo, M., Poole, R. K. et al. (2013) Bacterial and archaeal globins-A revised perspective, Biochim. Biophys. Acta. 1834, 1789-1800.
5. Gesto-Borroto, R., Sánchez-Sánchez, M. and Arredondo-Peter, R. (2015) A bioinformatics insight to rhizobial globins: gene identification and mapping, polypeptide sequence and phenetic analysis, and protein modeling., F1000Research. 4, 117.
6. Basak, P., Kundu, N., Pattanayak, R. et al. (2015) Denaturation properties and folding transition states of leghemoglobin and other heme proteins., Biochemistry (Moscow). 80, 463-472.
7. Codutti, L., Picotti, P., Marin, O., et al. (2009) Conformational stability of neuroglobin helix F – possible effects on the folding pathway within the globin family., FEBS J. 276, 5177-5190.
8. Culbertson, D. S. and Olson, J. S. (2010) Role of heme in the unfolding and assembly of myoglobin., Biochemistry. 49, 6052-6063.
9. Meinhold, D. W. and Wright, P. E. (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion., Proc. Natl. Acad. Sci. USA. 108, 9078-9083.
10. Mu, J., Li, L., Guo, Y., et al. (2010) Spectroscopic study on acid-induced unfolding and refolding of apo-neuroglobin., Spectrochim. Acta Part A. 75, 1600-1604.
11. Vahidi, S., Stocks, B. B., Liaghati-Mobarhan, Y. et al. (2013) Submillisecond protein folding events monitored by rapid mixing and mass spectrometry-based oxidative labeling., Anal. Chem. 85, 8618-8625.
12. Xu, M., Beresneva, O., Rosario, R. et al. (2012) Microsecond folding dynamics of apomyoglobin at acidic pH., J. Phys. Chem. B. 116, 7014-7025.
13. Kikuchi, T. (2002) Application to the prediction of structures and active sites of proteins and peptides. in Recent research developments in protein engineering. (Pandalai, S. G., ed) pp. 1-48., Research Signpost, Kerala.
14. Kikuchi, T., Némethy, G. and Scheraga, H. A. (1988) Prediction of the location of structural domains in globular proteins., J. Protein Chem. 7, 427–471.
15. Ichimaru, T. and Kikuchi, T. (2003) Analysis of the differences in the folding kinetics of structurally homologous proteins based on predictions of the gross features of residue contacts., Proteins: Struct. Funct. Genet. 51, 515-530.
16. Nishimura, C., Prytulla, S., Dyson, J. et al. (2000) Conservation of folding pathways in evolutionary distant globin sequences., Nature Struct. Biol. 7, 679-686.
17. Matsuoka, M., Fujita, A., Kawai, Y. et al. (2014) Similar structures to the E-to-H helix unit in the globin-like fold are found in other helical folds., Biomolecules. 4, 268-288.
18. Thompson, J. D., Gibson, T. J., Plewniak, F., et al. (1997) The clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools., Nucl. Acids Res. 24, 4876-4882.
19. Case, D. A., Darden, T. A., Cheatham, T. E., et al. (2012) AMBER 12, University of California, San Francisco.
20. Götz, A. W., Williamson, M. J., Xu, D., et al. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born., J. Chem. Theory Comput. 8, 1542-1555.
21. Salomon-Ferrer, R., Götz, A. W., Poole, D., et al. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. , J. Chem. Theory Comput. 9, 3878–3888.
22. Sindhikara, D. J., Yoshida, N. and Hirata, F. (2012) Placevent: An algorithm for prediction of explicit solvent atom distribution-application to HIV-1 protease and F-ATP synthase., J. Comput. Chem. 33, 1536-1543.
23. Kovalenko, A. and Hirata, F. (1998) Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. , Chem. Physics Lett. 290, 237-244.
24. Nakajima, S., Alvarez-Salgado, E., Kikuchi, T. et al. (2005) Prediction of folding pathway and kinetics among plant hemoglobins using an average distance map method., Proteins: Struc. Funct. Bioinf. 61, 500-506.
25. Eun, Y. J., Kurt, N., Sekhar, A. and Cavagnero, S. (2008) Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin., J. Mol. Biol. 376, 879-897.
26. Tarricone, C., Galizzi, A., Coda, A., et al. (1997) Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla* sp., Structure. 5, 497-507.
27. Milani, M., Pesce, A., Ouellet, Y., et al. (2004) Heme-ligand tunneling in group I truncated hemoglobins., J. Biol. Chem. 279, 21520-21525.

*****