Characterization of haemoglobin from Actinorhizal plants – An *in silico* approach

SANGHATI BHATTACHARYA[1], ARNAB SEN[2,*], SUBARNA THAKUR[2] and LOUIS S TISA[3]

[1]*Department of Botany, University of North Bengal, Raja Rammohunpur, Siliguri, India*

[2]*Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri 734013, India*

[3]*Department of Cellular, Molecular and Biomedical Sciences, University of New Hampshire, Durham, NH, USA*

**Corresponding author (Email, senarnab_nbu@hotmail.com)*

Plant haemoglobins (Hbs), found in both symbiotic and non-symbiotic plants, are heme proteins and members of the globin superfamily. Hb genes of actinorhizal Fagales mostly belong to the non-symbiotic type of haemoglobin; however, along with the non-symbiotic Hb, *Casuarina* sp. posses a symbiotic one (symCgHb), which is expressed specifically in infected cells of nodules. A thorough sequence analysis of 26 plant Hb proteins, currently available in public domain, revealed a consensus motif of 29 amino acids. This motif is present in all the members of symbiotic class II Hbs including symCgHb and non-symbiotic Class II Hbs, but is totally absent in Class I symbiotic and non-symbiotic Hbs. Further, we constructed 3D structures of Hb proteins from *Alnus* and *Casuarina* through homology modelling and peeped into their structural properties. Structure-based studies revealed that the *Casuarina* symbiotic haemoglobin protein shows distinct stereochemical properties from that of the other *Casuarina* and *Alnus* Hb proteins. It also showed considerable structural similarities with leghemoglobin structure from yellow lupin (pdb id 1GDI). Therefore, sequence and structure analyses point to the fact that symCgHb protein shows significant resemblance to symbiotic haemoglobin found in legumes and may thus eventually play a similar role in shielding the nitrogenase from oxygen as seen in the case of leghemoglobin.

[Bhattacharya S, Sen A, Thakur S and Tisa LS 2013 Characterization of haemoglobin from Actinorhizal plants – An *in silico* approach. *J. Biosci.* **38** 777–787] **DOI** 10.1007/s12038-013-9357-0

1. Introduction

Haemoglobins (Hbs) are heme proteins with distinguished roles in oxygen transport and respiration in animals. They are found ubiquitously in eukaryotes and in many bacteria (Dordas *et al.* 2003). Plant Hbs, structurally similar to animal Hbs and myoglobins, were first characterized from the root nodules of leguminous plants (Kubo 1939). Initially, plant Hb proteins were thought to be restricted to plant species carrying out symbiotic nitrogen fixation, but further analysis revealed their presence in non-nodulating plants (Arredondo-Peter *et al.* 1997). Symbiotic Hbs (s-Hbs) are found mostly in the root nodules of leguminous plants, while non-symbiotic Hbs (ns-Hbs) are expressed in both leguminous and non-leguminous plants. In plants, three distinct types of haemoglobins have been characterized: symbiotic (s-Hb), nonsymbiotic (ns-Hb) and truncated (t-Hb) haemoglobins (Duff *et al.* 1997; Arredondo-Peter *et al.* 1998). s-HBs serves to facilitate oxygen diffusion to the bacterial endosymbiont and to buffer the free oxygen concentration at a low tension to protect the nitrogenase from oxygen-inactivation (Appleby 1992). ns-Hbs are mainly involved in NO scavenging (Gupta *et al.* 2011) and their expression is directly associated with protection

Keywords. Actinorhizal plant; haemoglobin (Hb); homology modelling; vesicle

Abbreviations used: CG, conjugate gradient; non-symCgHb, non-symbiotic haemoglobin of *Casuarina glauca*; ns-Hb, nonsymbiotic haemoglobins; s-Hb, symbiotic haemoglobins; symCgHb, symbiotic haemoglobin of *Casuarina glauca*; t-Hb, truncated haemoglobins

Published online: 1 October 2013

against hypoxic challenge (Hunt *et al.* 2002). They are further categorized into two classes: class I, with dramatically different oxygen-binding properties compared to s-Hbs, and class II, with similar oxygen-binding properties to s-Hbs (Dordas *et al.* 2003). t-Hbs share some characteristics with ns-Hbs, but the exact role of t-Hbs remains largely unknown (Watts *et al.* 2001). Amongst actinorhizal plants, *Casuarina glauca* contain both symbiotic (symCgHb) and non-symbiotic haemoglobin (non-symCgHb). In fact, it is the only actinorhizal species that is known to contain s-Hb, and is expressed at a high level in the nodules that forms a nitrogen-fixing symbiosis with *Frankia.* The symCgHb protein is similar to other nodule-expressed sym-Hb proteins in respect to having an unusually low pH-sensitive oxygen off-rate (Dordas *et al.* 2003).

In the present study, we have carried out a thorough *in silico* sequence-based study to identify different trends in the physical and chemical properties of haemoglobin proteins from different plants. Analysis of physiochemical parameters revealed that class II ns-Hb proteins shared some features with class II s-Hb proteins. A characteristic motif of 29 amino acid residues, present in all of the class II s-Hb and class II ns-Hb proteins, was also identified. Further, 3D structures of Hb proteins from *Alnus* and *Casuarina* (both non-symCgHb and symCgHb) have been derived by homology modelling approaches and have been analysed to explore the structural features that influence the functional differences of these proteins.

2. Materials and methods

2.1 *Sequences retrieval*

Sequences of symbiotic and non-symbiotic types of Hb genes were retrieved from NCBI (*http://www.ncbi.nlm.nih.gov/*) database. Accession number and the details of retrieved sequences are listed in table 1.

2.2 *Physiochemical parameter analysis*

Physiochemical data were generated from the ProtParam (*http://web.expasy.org/protparam/*) software using ExPASy server.

2.3 *Identification of domains and motifs*

The amino acid sequences were aligned and subjected to BLOCK MAKER for domain analysis. Accordingly, the blocks were fed to MEME suit for motif elicitation followed by MAST (*http://meme.ebi.edu.au/meme/cgi-bin/meme.cgi*) search. A BLAST search was carried out on the screened motifs for identification of conserved protein motifs.

2.4 *Constructions of 3D models*

Three-dimensional protein models of *Alnus* and *C. glauca* ns-Hbs along with s-Hbs were constructed by homology modelling approach. A suitable template was identified by PSI-BLAST (Altschul *et al.* 1997) against the Protein Data Bank (PDB) proteins available at the NCBI web server and an appropriate template was selected on the basis of sequence and phylogenetic similarity (Centeno *et al.* 2005). Multiple sequence alignments were performed using ClustalW2 (Thompson *et al.* 1994). Three-dimensional models of the proteins were constructed by using the MODELLER 9v6 program (Sali and Blundell 2009) based on its alignment with the template protein (Centeno *et al.* 2005). The constructed model was subjected to energy minimization for the refinement of the structures, and keeping a harmonic constraint of 100 kJ/mol/Å^2. The steepest descent (SD) and conjugate gradient (CG) methods (both 100 steps) were used to remove existing bad sectors between the protein atoms and protein structure geometry. The refined models were submitted to ProFunc (*http://www.ebi.ac.uk/thronton-srv/databases/ProFunc*) (Laskowski *et al.* 2005) to garner information about functionally important regions of the protein.

2.5 *Evaluation of refined models*

Refinements of the modelled structures were performed using a variety of web-based servers. A stringent refinement policy was adopted in order to ensure that the modelled structures were void of any structural errors. To ensure accuracy and reliability, the refined protein models were evaluated via ProSA (Wiederstein and Sippl 2007) and VERIFY3D (Eisenberg *et al.* 1997). Ramachandran plot (Ramachandran *et al.* 1963) was used to assess the constructed model for its backbone conformation and also to inspect the favourable and unfavourable regions of the modelled structure. We used ERRAT (Colovos and Yeates 1993) (*http://nihserver.mbi.ucla.edu/ERRATv2*) and SAVES (Structure analysis and verification server) (*http://nihserver.mbi.ucla.edu/SAVES/*) for other verifications.

2.6 *Studying intrinsic dynamics of the protein model*

The structural dynamics task was accomplished using various web-based strategies. WEBnm (*http://www.bioinfo.no/tools/normalmodes*) (Hollup *et al.* 2005) was used to calculate the slowest modes and related deformation energies. ElNemo (*http://igs-server.cnrs-mrs.fr/elnemo/index.html*) (Suhre and Sanejouand 2004) was utilized to calculate the normal mode analysis of the proteins that contribute to the corresponding protein movements. Normal mode analyses

Table 1. Accession numbers of the symbiotic and non-symbiotic types of Hb protein sequences

Sample no.	Source organisms	Types	Accession No.
1	*Parasponia andersonii*	Symbiotic class I	1212354A
2	*Casuarina glauca*	Symbiotic class II	P08054.2
3	*Lupinus luteus*	Symbiotic class II	AAC04853.1
4	*Sesbania rostrata*	Symbiotic class II	CAA31859.1
5	*Vigna unguiculata*	Symbiotic class II	AAA86756.1
6	*Phaseolus vulgaris*	Symbiotic class II	AAA33767.1
7	*Glycine max*	Symbiotic class II	CAA23729.1
8	*Medicago sativa*	Symbiotic class II	AAB48005.1
9	*Vicia faba*	Symbiotic class II	CAA90869.1
10	*Pisum sativum*	Symbiotic class II	BAA31155.1
11	*Lotus japonicus*	Symbiotic class II	BAE46738.1
12	*Casuarina glauca*	Non-symbiotic class I	P23244.1
13	*Alnus firma*	Non-symbiotic class I	BAE75956.1
14	*Glycine max*	Non-symbiotic class I	AAA97887.1
15	*Medicago sativa*	Non-symbiotic class I	AAG29748.1
16	*Gossypium hirsutum*	Non-symbiotic class I	AAX86687.1
17	*Lotus japonicus*	Non-symbiotic class I	BAE46739.1
18	*Arabidopsis thaliana*	Non-symbiotic class I	AEC06463.1
19	*Trema tomentosa*	Non-symbiotic class I	CAA68405.1
20	*Oryza sativa*	Non-symbiotic class I	AAK72231.1
21	*Zea mays*	Non-symbiotic class I	NP_001104966.1
22	*Hordeum vulgare*	Non-symbiotic class I	AAB70097.1
23	*Brassica napus*	Non-symbiotic class II	AAK07741.1
24	*Gossypium hirsutum*	Non-symbiotic class II	AAK21604.1
25	*Cichorium intybus*	Non-symbiotic class II	CAA07547.1
26	*Arabidopsis thaliana*	Non-symbiotic class II	AEE74919.1

predict the probable movements of the proteins and aid in selection of the slowest activity of proteins (Hollup *et al.* 2005). Determination of the lowest frequency modes was performed using MolMovDB (*http://molmovdb.org/*) (Alexandrov *et al.* 2005). Solvent accessibility of the amino acid residues in the modelled proteins was determined by using ASA-view (*http://gibk26.bse.kyutech.ac.jp/~shandar/netasa/asaview/*) (Ahmad *et al.* 2004).

3. Results and discussion

3.1 *Physiochemical parameter analysis*

Physiochemical features of haemoglobin protein sequences from various plants are provided in table 2. The total number of amino acid residues ranged from 145 to 167 with variable molecular weight. The pI values of class II s-Hb ranged from 5.29 to 6.97, while class II ns-Hb ranged from 5.4 to 5.89. The pI values for only known class I s-Hb from *Parasponia* was found to be 8.59 (Wittenberg et al. 1986). In case of class I ns-Hb, the pI value ranged from 7.84 to 9.3. The low pI value of Class II s-Hb and ns-Hb can be attributed to the dominance of surface metal-OH species (Satoshi and Makoto 2005) and explains the higher oxygen binding capacity of metal cofactors of class II s-Hb and ns-Hb members relative to class I members. Results also reveal that surface of class II ns-Hb and s-Hb proteins are rich in negatively charged residues, while class I ns-Hb and s-Hb proteins contain more positively charges residues on their surface.

The *in vivo* half-life of a protein is calculated via the instability index (Guruprasad *et al* 1990), which indicates the extent of stability of the proteins. Previously, it was reported that proteins having instability index of more than 40 have an *in vivo* half-life of less than 5 h, while those proteins having instability index value less than 40 have a longer *in vivo* half-life of 16 h (Rogers *et al.* 1986). In our study, the instability index values of most of the studied haemoglobin protein sequences were found to be lower than 40, except for class II ns-Hb from *Arabidopsis thaliana*. The thermostability of the proteins was assessed by the aliphatic index. The aliphatic index is directly proportional to the thermostability, and is defined as relative volume occupied

Table 2. Physiochemical features of symbiotic and non-symbiotic types of Hb protein sequences from various plants

Organisms	No of amino acids	Molecular weight	Theoritical pI	Total no. negative residues	Total no. of positive residues	Instability index	Aliphatic index	GRAVY
Parasponia andersonii(S-I)	162	18.18	8.59	19	21	38.32	86.05	−0.098
Pisum sativum(S-II)	146	15.94	5.57	19	15	29.85	92.88	0.019
Vicia faba(S-II)	146	15.85	5.92	18	16	33.24	98.22	0.085
Medicago sativa(S-II)	147	15.93	5.33	18	14	36.41	86.26	−0.012
Lotus japonicus(S-II)	147	15.75	5.29	18	13	19.45	87.69	−0.007
Glycine max(S-II)	151	16.26	6.07	18	16	22.94	95.7	0.055
Phaseolus vulgaris(S-II)	146	15.62	6.09	15	14	21.08	95.07	0.009
Vigna unguiculata(S-II)	145	15.36	6.11	17	16	14	97.03	0.108
Sesbania rostrata(S-II)	148	15.9	5.61	19	16	18.47	92.97	−0.007
Lupinus luteus(S-II)	154	16.75	5.79	19	16	19.34	101.95	0.066
Casuarina glauca(S-II)	152	17.24	6.97	20	20	32.91	91.91	−0.37
Cichorium intybus(ns-II)	161	18.02	5.51	25	21	28.61	90.87	−0.188
Gossypium hirsutum(ns-II)	159	18.1	5.44	28	21	36.67	82.2	−0.473
Arabidopsis thaliana(ns-II)	158	17.87	5.4	26	19	41.4	92.59	−0.312
Brassica napus(ns-II)	161	18.32	5.89	26	22	30.63	87.2	−0.363
Hordeum vulgare(ns-I)	162	18.04	7.84	21	22	27.97	86.23	−0.096
Zea mays(ns-I)	165	18.28	6.32	23	22	22.38	88.24	−0.016
Oryza sativa(ns-I)	167	18.61	9.3	18	22	21.03	83.05	−0.093
Trema tomentosa(ns-I)	161	18.15	8.59	20	22	31.76	82.98	−0.22
Arabidopsis thaliana(ns-I)	160	18.03	8.46	20	22	32.47	85.31	−0.148
Lotus japonicus(ns-I)	161	18.04	9	18	22	33.4	80.56	−0.201
Gossypium hirsutum(ns-I)	163	18.38	8.77	21	24	20.56	86.2	−0.08
Medicago sativa(ns-I)	160	17.96	9.08	19	23	23.79	81.69	−0.255
Glycine max(ns-I)	161	18.05	8.97	19	22	27.05	79.94	−0.245
Alnus firma(ns-I)	160	17.9	8.95	19	22	26.72	84.12	−0.156
Casuarina glauca(ns-I)	160	17.85	8.93	20	23	24.18	81.12	−0.22

by the aliphatic side chain (Ikai 1980). For *Parasponia* Hb, the aliphatic index was found to be 86.05. With class II s-Hb the values ranged from 86.26-101.95, while the values for class I and class II ns-Hbs ranged from 79.94 to 88.24 and 82.2 to 92.59, respectively. Thus, the class II members of both s-Hb and ns-Hb were more thermostable than class I members based on the higher values of aliphatic index. GRAVY (grand average of hydropathicity) values reflect the hydrophobicity of the amino acids. An increase in GRAVY values indicates that a protein tend to be more hydrophobic in nature (Klein and Thongboonkerd 2004). GRAVY values were found to be highest for class II s-Hb proteins, followed by class II ns-Hb members. Class I members were found to possess comparatively low GRAVY scores. Investigation of various physiochemical parameters revealed that class II members of both s-Hb and ns-Hb possess similar characteristics, which are quite discrete from those of the class I members.

Table 3. Distinct motif identified in various symbiotic and non-symbiotic types of plant Hb protein sequences

Types	Motif No.	Motif length	Motif sequence
Class II (symbiotic And non-symbiotic)	1	29	PQNNPKLQAHAEKVFGMTCDSAIQLRANG
All	1	50	CFTEEQEALVVKSWEVMKQNIPNYGLRFYTKIFEIAPSARNMFSFLRDSN
	2	49	HFQYGVVDPHFEVTKFALLRTIKEAVPDMWSPEMMNAWWQAYDQLVAAI

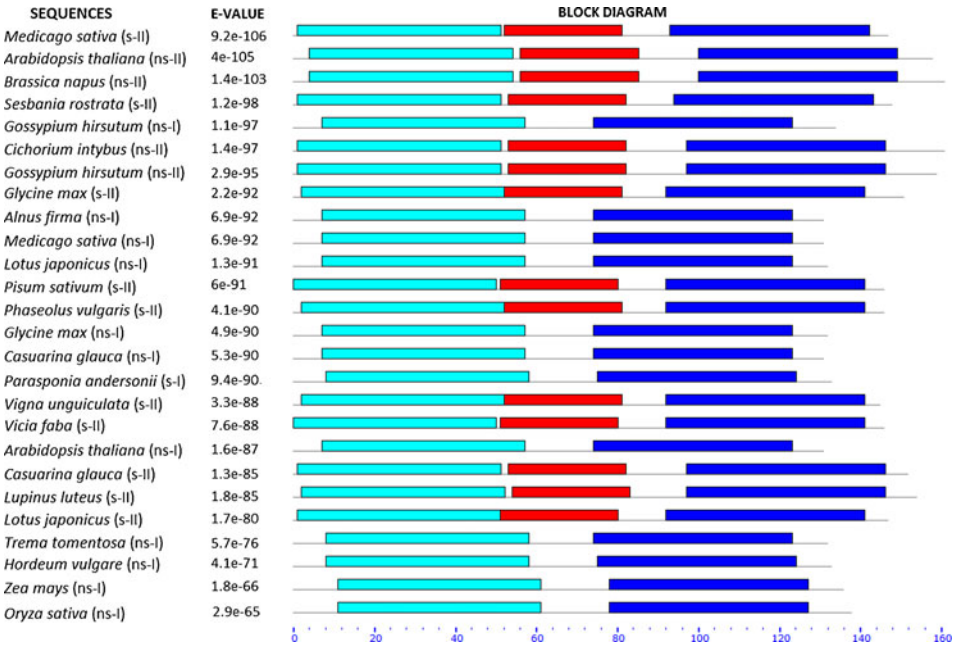

Figure 1. Different motifs identified in various plants Hb protein sequences.

3.2 *Motif finding*

It is evident from the results of multiple sequence alignments that there is a distinct homology between plant class II s-Hb and ns-Hb sequences. A total of seven conserved motifs were observed and some short-length conserved regions were also found. The seven recognized motifs of varying width were subjected to protein BLAST for confirmation of the motif annotations. Results revealed that only three out of seven motifs had similarities with GLOBIN family. One out of the three motifs had a distinct 29 amino acid stretch (PQNNPKLQAHAEKVFGMTCDSAIQLRANG). This motif was found in class II ns-Hb and s-Hb proteins, but was totally absent in class I Hb proteins. The symCgHb also possesses this unique motif. The distinct motif of 29 amino acid residues further points towards the functional similarity of class II ns-Hb proteins with that of s-Hb. We also found other motifs of approximately ~49 amino acid and ~50 amino acid residues for all of the Hb protein sequences. table 3 shows the motifs of plant Hb protein sequences from

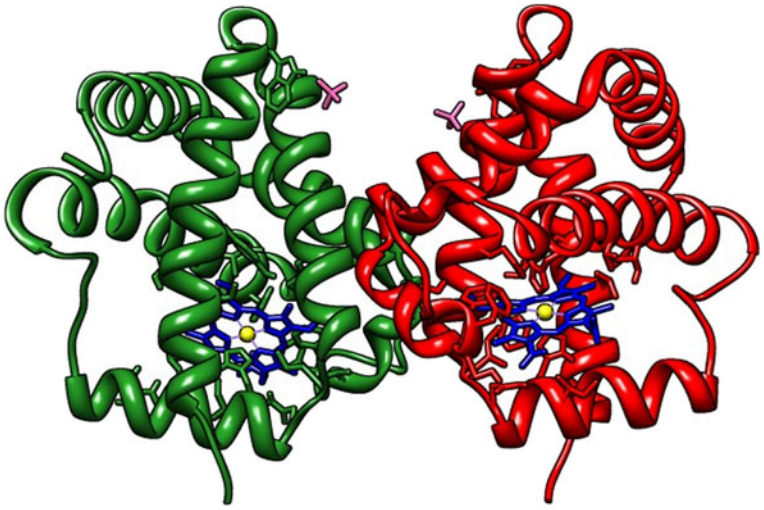

Figure 2. Three-dimensional structure of non-symbiotic *Casuarina glauca* Hb protein constructed by homology modelling technique.

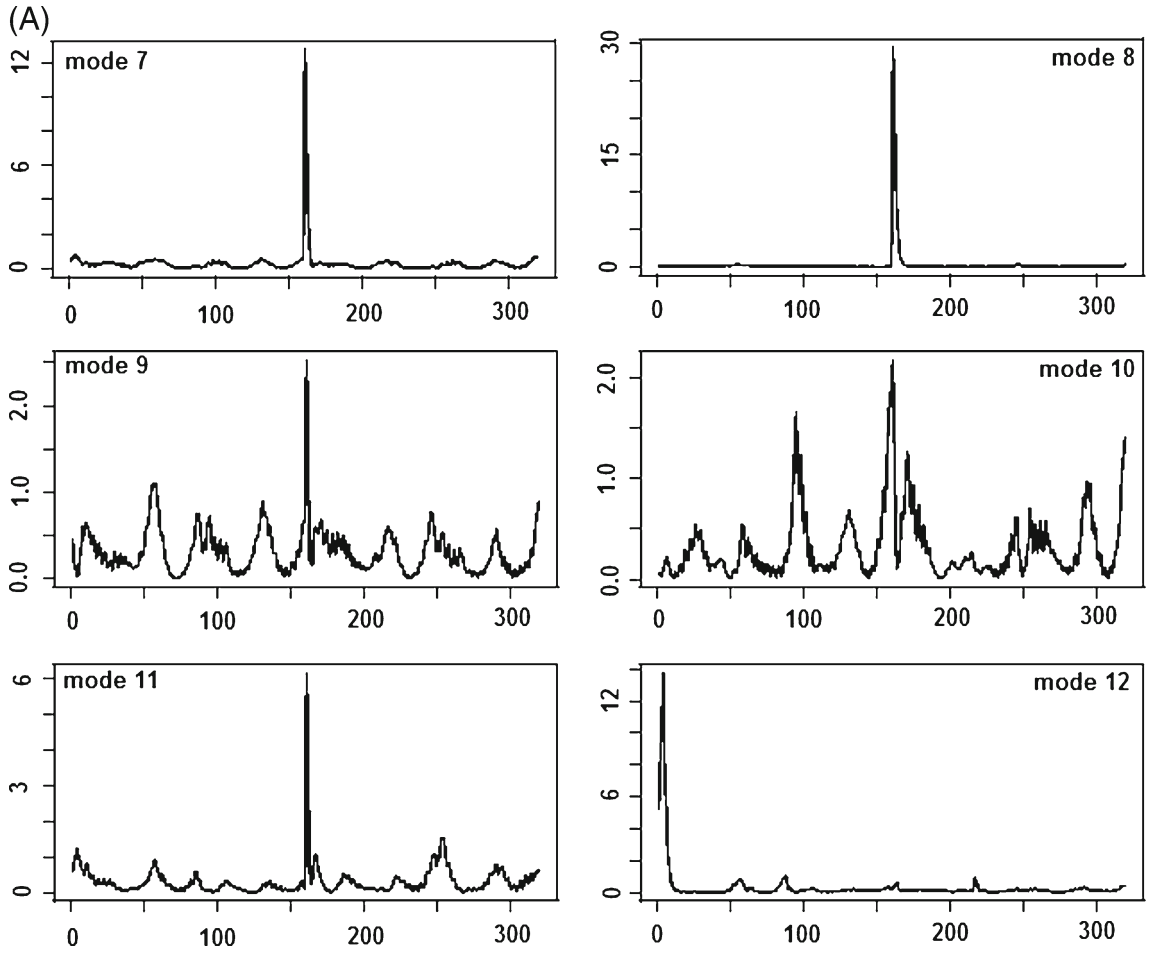

Figure 3. Normalized atomic displacement plot calculated for modes 7 to 12 in Hb proteins of (**A**) *Alnus* and (**B**) Ns-*Casuarina*. The figure shows the plot of the normalized square atomic displacements which represents the square of the displacement of each C-alpha atom. The highest values corresponded to the most displaced regions and residues with maximum displacements associated with functional sites. X and Y-axis denote residue index in sequence and normal mode of square atomic displacement respectively.

various actinorhizal plants. The position of these motifs in the various protein sequences are provided in figure 1.

3.3 *Three-dimensional models of Hb proteins*

To infer about the structural details of the Hb from actinorhizal plants, the 3D structures of the proteins were predicted by homology modelling technique. The PSI-BLAST search found the crystal structure of the *Trema tomentosa* Hb protein [3QQQ - PDB ID] to be the best template for *A. firma* and *C. glauca* class I ns-Hb proteins. *T. tomentosa* ns-Hb is a homodimeric protein and this template had a-match with an e-value of 2e−90 and 2e−91 with *A. firma* and *C. glauca* ns-Hb proteins, respectively. The *Trema* Hb protein also had 83% and 84% sequence similarities with the *A. firma* and *C. glauca* Hb proteins, respectively. The haemoglobin protein from *Lupinus luteus* [PDB ID-1GDI] had an e-value of 2e−47 and 54% similarity to symCgHb and was found to be the best template for this protein. The modelled structures of *Alnus* and *Casuarina* Hb proteins were also homodimers. Each subunit of modelled proteins has a hetero atom of heme group containing Fe at the center, which is the core functional region of the protein. Figure 2 displays the modelled structure of the non-symCgHb protein. Functional analysis of the Hb-proteins revealed the presence of nests (Watson and Milner-White 2002) in each chain. These nests are structurally important motifs found in functionally important regions. The non-symCgHb protein showed 3 nests in its structure, whereas its counterpart symCgHb showed 1 nest. CASTp analysis revealed the presence of pockets for ligand interactions on the

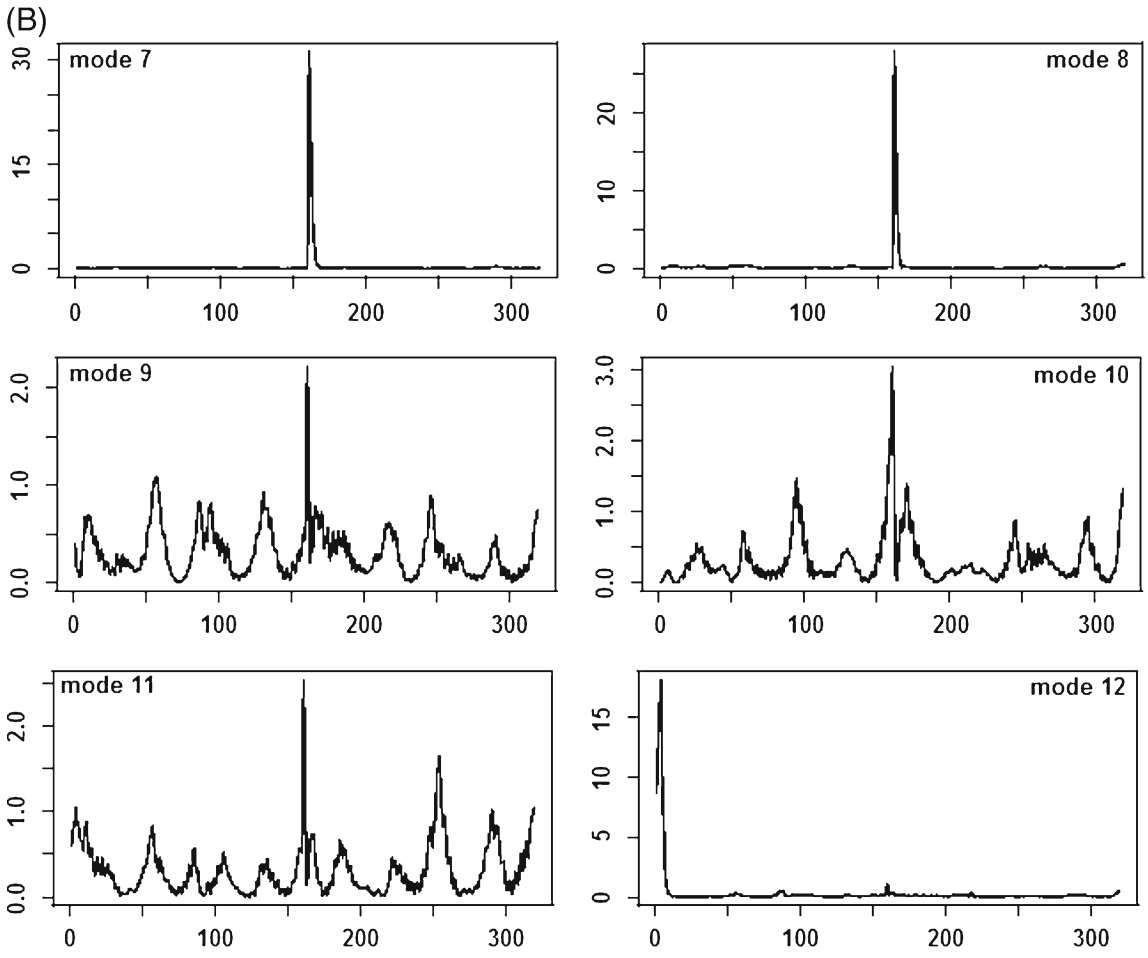

Figure 3. continued.

protein surface. *Alnus* and *Casuarina* ns-Hb proteins showed 23 and 29 pockets, respectively, whereas s-Hb displayed 22 pockets. No DNA binding templates and enzyme active site templates were recognized in the protein structures.

The constructed models were further assessed to estimate their accuracy. The z-score obtained from ProSA analysis specifies the overall quality of the models and determines the extent to which total energy of the modelled structure drifts from energy distributions of the random conformations (Wiederstein and Sippl 2007). The *Alnus* Hb, non-symCgHb and symCgHb proteins had z scores of −7.35, −7.72 and −7.01, respectively. These results specify that our models are very much within the range of scores normally found for proteins of comparable size and the outcome of the energy plot signifies that 3D models of Hb proteins are reliable and precise. VERIFY 3D analysis revealed that 80.063% of the residues had an average 3D-1D score >0.2. The Ramachandran plot for the *Alnus* and *Casuarina* ns-Hb proteins revealed that 95.5% of the total residues were present in the most favored regions, while symCgHb exhibited 92.5%. A good quality model is expected to have more than 90% in the most favoured regions of the Ramachandran plot (Ramachandran *et al.* 1963). PROVE, VERIFY 3D and ERRAT results for all three proteins illustrated that the overall quality of the models are good. These results imply that the stereochemical properties and quality of modelled structures are quite consistent.

In normal mode analysis (NMA), the first six modes matching with global rotation and translation of the system are generally ignored (Hollup *et al.* 2005) and hence the lowest frequency mode of concern is the seventh one. NMA of the Hb protein showed that low deformation energies were associated with relatively rigid regions in the protein. From figures 3 and 4, it can easily be recognized that normalized atomic displacement plot of symCgHb

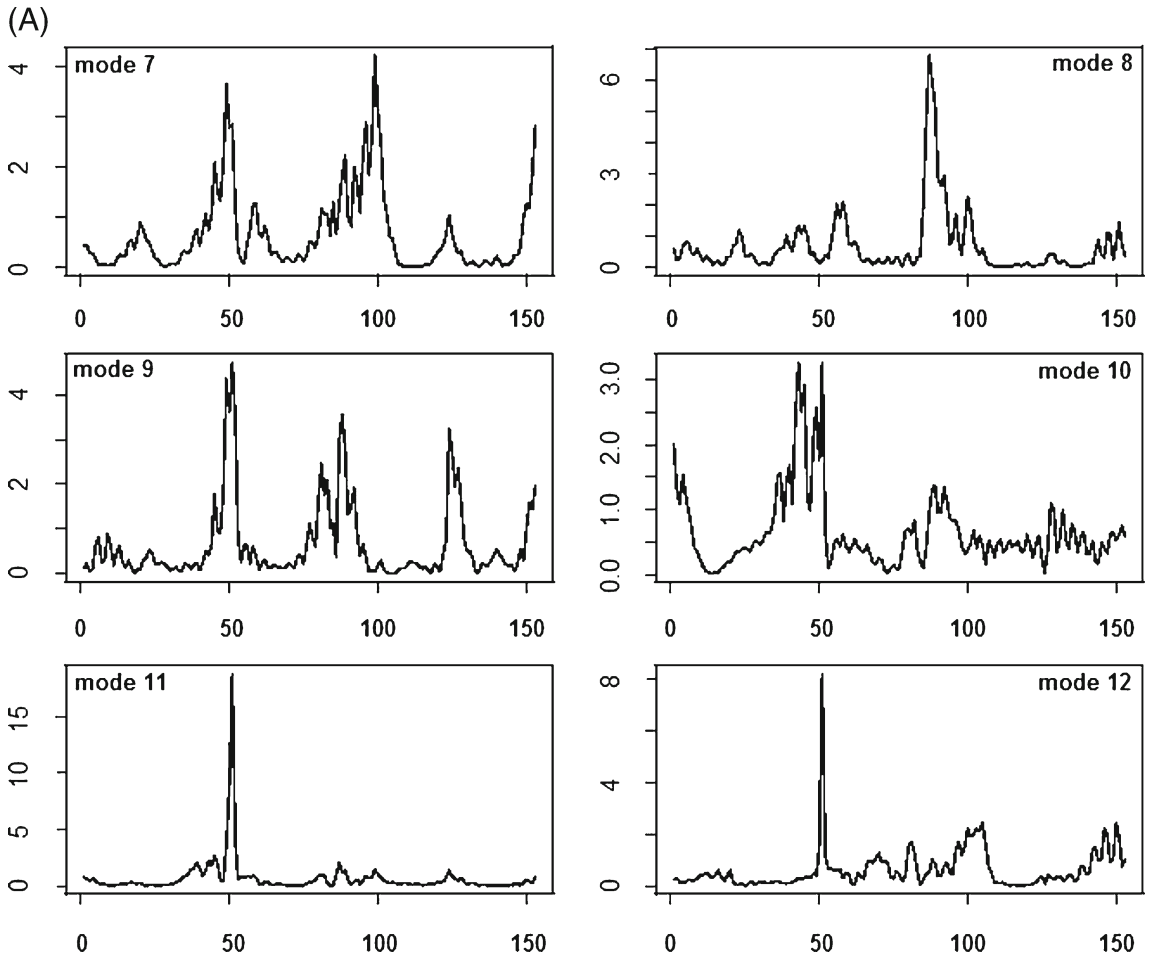

Figure 4. Normalized atomic displacement plot calculated for modes 7 to 12 in Hb proteins of (**A**) *Lupin* abd (**B**) s-*Casuarina*. The figure shows the plot of the normalized square atomic displacements which represents the square of the displacement of each C-alpha atom. The highest values corresponded to the most displaced regions and residues with maximum displacements associated with functional sites. X and Y-axis denote residue index in sequence and normal mode of square atomic displacement respectively.

shows dissimilarities with non-symCgHb from the seventh mode onwards. NMA indicated the vibrational and thermal properties of the protein at the atomic level. *Alnus* and *Casuarina* ns-Hb proteins had the lowest deformation energies of 2699.93 and 2700.92. The non-symCgHb protein showed higher deformation energy of 3869.35, which is significantly diverse from ns-Hb proteins in the seventh mode. B factors calculated from ElNemo analysis were based on the first 100 normal modes and were scaled to match the overall B factors. In the case of *Alnus* and *Casuarina* Hb proteins, very low negative correlations were obtained between the computed and observed B factors. The deformation energies implied that the seventh mode of non-symCgHb had comparatively large rigid regions than symCgHb. The B factors calculated from ElNemo analysis signifies that the models of *Alnus* and *Casuarina* ns-Hb proteins contain enough rigid regions and are less flexible, while the symCgHb protein model was more flexible. ASAVIEW analysis of the solvent accessibility for the modelled proteins pointed out that the accessible residues were present on the outermost ring of the spiral. Figure 5 shows the solvent accessibility plot of the symCgHb protein. The majority of negatively charged residues and polar uncharged residues were present on the outermost surface and hydrophobic residues were confined to the inner rings of the spiral. However, ns-Hb had positively charged residues predominantly present on the outermost surface (data not shown).

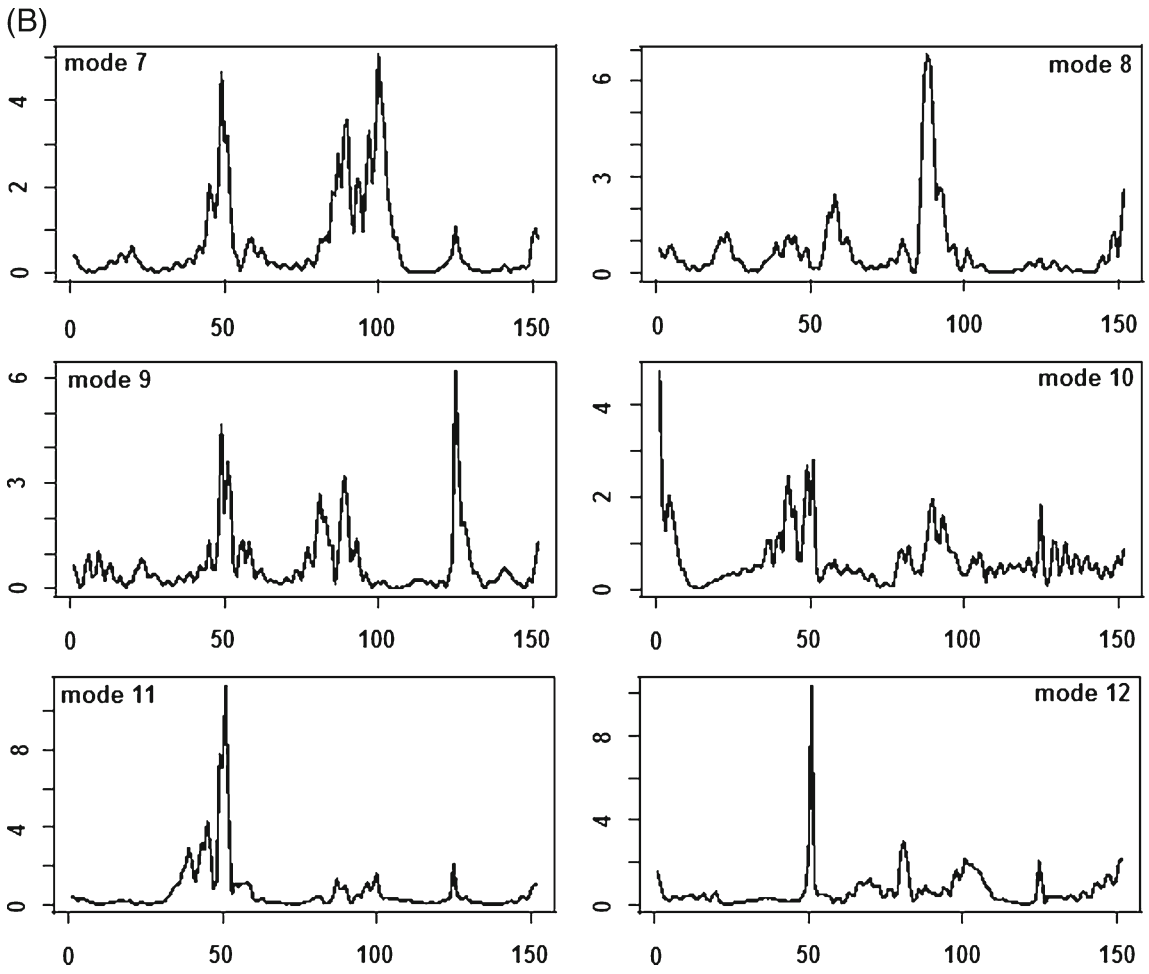

Figure 4. continued.

4. Conclusion

An investigation of the physiochemical parameters, including theoretical pI, aliphatic index and GRAVY, found that class II s-Hb and ns-Hb proteins share many features, which set them apart from those belonging to class I. Members of class II haemoglobin also have sequence-based similarity and were found to share a common motif consisting of 29 amino acid sequences. The symCgHb was also found to bear this characteristic motif, which is totally absent in non-symCgHb. This distinctive sequence features could further be utilized for designing strategy for cloning the putative genes based on PCR amplification using degenerate primers

The 3D structures presented here are the first ever reported models of *Alnus* and *Casuarina* Hb proteins. The protein models are homo-dimeric with the heme group controlling the conformational reaction of the protein. Clefts and cavities present of the protein surface represents the active sites, which are responsible for the inherent physicochemical features, and have a vital role in protein functioning. NMA of the Hb protein demonstrated that low deformation energies are associated with relatively rigid regions in the protein and revealed that symCgHb and non-symCgHb have different motional properties.

The combined results from sequence analysis, motif analysis and comparison of 3D protein models indicate that symCgHb and non-symCgHb have distinctively different properties. In many aspects, the symCgHb protein is similar to the nodule-expressed symbiotic haemoglobin of legumes. Thus, it may play a similar role in nitrogenase protection mechanism and provide a plausible explanation towards the absence of vesicle (the site for N_2 fixation in *Frankia*) in *Frankia–Casuarina* symbiosis.

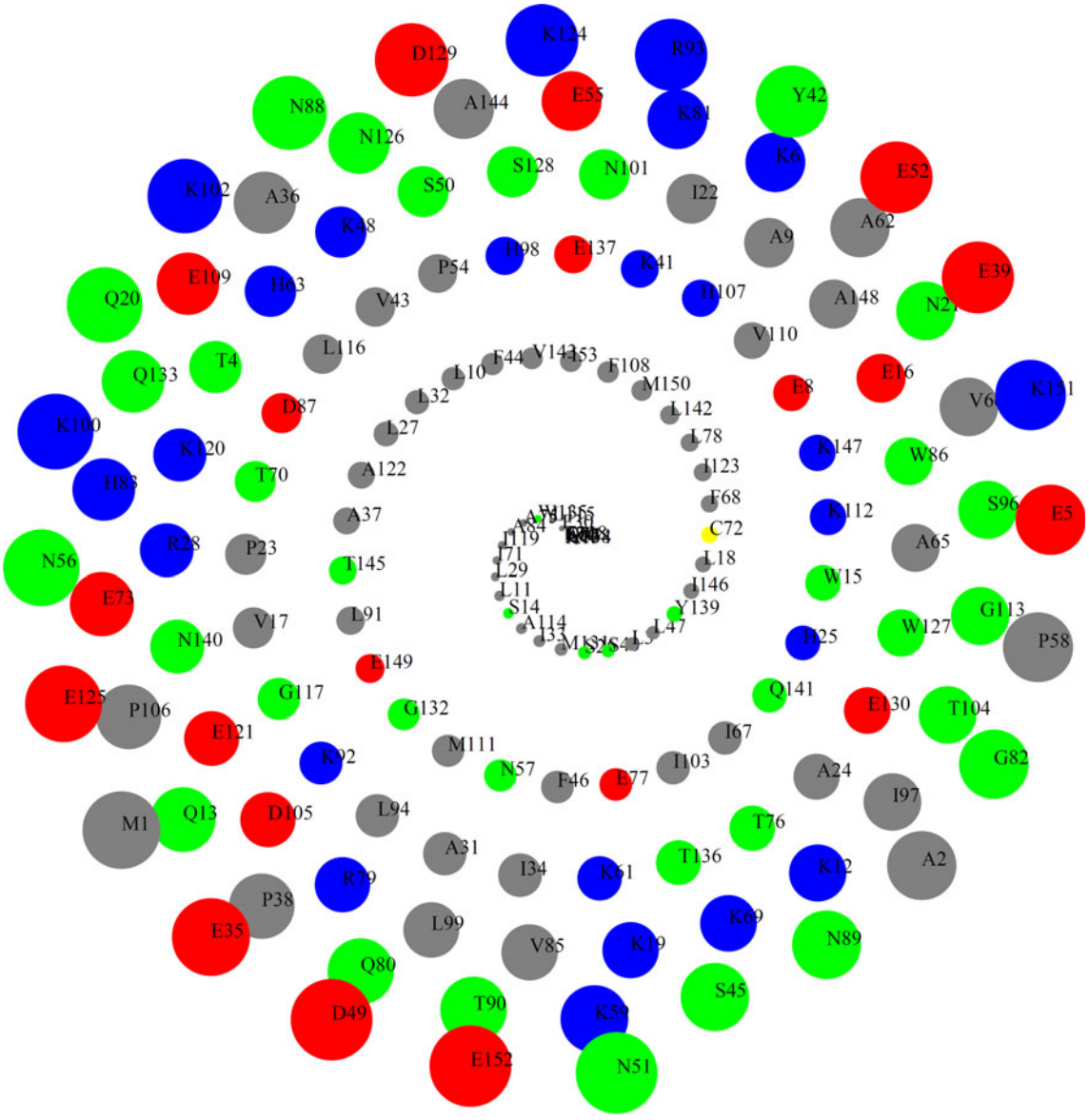

Figure 5. Solvent accessibility plot of the s-*Casuarina* Hb protein.

References

Ahmad S, Gromiha M, Fawareh H and Akinori S 2004 ASAview: database and tools for solvent accessibility representation in proteins. *BMC Bioinform*. **5** 51

Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D and Gerstein M 2005 Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. *Protein Sci*. **14** 633–643

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25** 3389–3402

Appleby CA 1992 The origin and functions of haemoglobin in plants. *Sci. Prog*. **76** 365–398

Arredondo-Peter R, Hargrove MS, Moran JF, Sarath G, Moran JF, Lohrman J, Olson JS and Klucas RV 1997 Rice hemoglobins. *Plant Physiol*. **115** 1259–1266

Arredondo-Peter R, Hargrove MS, Moran JF, Sarath G and Klucas RV 1998 Plant hemoglobins. *Plant Physiol*. **118** 1121–1126

Centeno NB, Planas-Iglesias J and Oliva B 2005 Comparative modeling of protein structure and its impact on microbial cell factories. *Microbial Cell Factories* **4** 20

Colovos C and Yeates TO 1993 Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. **2** 1511–1519

Dordas C, Rivoal J and Hill RD 2003 Plant haemoglobins, nitric oxide and hypoxic stress. *Ann. Bot.* **91** 172–178

Duff SMG, Wittenberg JB and Hill RD 1997 Expression, purification, and properties of recombinant barley (*Hordeum* sp.) hemoglobin. *J. Biol. Chem.* **272** 16746–16752

Eisenberg D, Luthy R and Bowie JU 1997 VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*. **277** 396–404

Gupta KJ, Hebelstrup KH, Mur LAJ and Igamberdiev AU 2011 Plant hemoglobins: important players at the crossroads between oxygen and nitric oxide. *FEBS Lett*. **585** 3843–3849

Guruprasad K, Reddy BVP and Pandit MW 1990 Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot. Eng.* **4** 155–164

Hollup SM, Salensminde G and Reuter N 2005 WEBnm@: a web application for normal mode analysis of proteins. *BMC Bioinform*. **6** 1–8

Hunt PW, Klok EJ, Trevaskis B, Watts RA, Ellis MH, Peacock WJ and Dennis ES 2002 Increased level of hemoglobin 1 enhances survival of hypoxic stress and promotes early growth in *Arabidopsis thaliana*. *Proc. Nat. Acad. Sci. USA* **99** 17197–17202

Ikai A 1980 Thermostability and aliphatic index of globular proteins. *J. Biochem.* **6** 1895–1898

Klein JB and Thongboonkerd V 2004 Overview of proteomics. *Proteomics Nephrol*. **141** 1–10

Kubo H 1939 Uber hamaprotein as den wurzelknollchen von leguminosen. *Acta Phytochim*. **11** 195–200

Laskowski RA, Watson JD and Thronton JM 2005 ProFunc: a server for predicting protein function from structure. *Nucleic Acids Res*. **33** 89–93

Ramachandran GN, Ramakrishnan C and Sasisekharan V 1963 Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7** 95–99

Rogers S, Wells R and Rechsteiner M 1986 Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234** 364–368

Sali A and Blundell TL 2009 Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234** 283–291

Satoshi T and Makoto F 2005 Role of surface OH groups in surface chemical properties of metal oxide films. *Mater. Sci. Engr.* **119** 265–267

Suhre K and Sanejouand YH 2004 ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*. **32** 610–614

Thompson JD, Higgins DG and Gibson T 1994 CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22** 4673–4680

Watson JD and Milner-White EJ 2002 The conformations of polypeptide chains where the main-chain parts of the successive residues are enantiomeric. Their occurrence in cation and anion-binding regions of proteins. *J. Mol. Biol.* **315** 187–198

Watts RA, Hunt PW, Hvitved AN, Hargrove MS, Peacock WJ and Dennis ES 2001 A hemoglobin from plants homologous to truncated hemoglobins of microorganisms. *Proc. Nat. Acad. Sci. USA* **98** 10119–10124

Wittenberg JB, Wittenberg BA, Gibson QH, Trinick MJ and Appleby CA 1986 The kinetics of the reactions of *Parasponia andersonii* hemoglobin with oxygen, carbon monoxide, and nitric oxide. *J. Biol. Chem.* **261** 13624–13631

Wiederstein M and Sippl MJ 2007 ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. **35** 407–410