

### Evolutionary trace analysis of plant haemoglobins: implications for site-directed mutagenesis

Duvvuru muni Rajasekhara Reddy\*

Division of biology, University of Fribourg, Fribourg, Switzerland;

Email: raja.duvvuru@unifr.ch; \* Corresponding author

received January 05, 2007; revised March 09, 2007; accepted March 20, 2006; published online March 27, 2007

#### Abstract:

Haemoglobins are found ubiquitously in eukaryotes and many bacteria. In plants, haemoglobins were first identified in species, which can fix nitrogen via symbiosis with bacteria. Recent findings suggest that another class of haemoglobins termed as non-symbiotic haemoglobins are present through out the plant kingdom and are expressed differentially during plant development. Limited data available suggests that non-symbiotic haemoglobins are involved in hypoxic stress and oversupply of nutrients. Due to lack of information on structurally conserved, functionally important residues in non-symbiotic haemoglobins, further studies to elucidate the molecular mechanisms underlying the biological role are hampered. To determine functionally important residues in non-symbiotic haemoglobins, I have analyzed a number of sequences from plant haemoglobin family, in the context of the known crystal structures of plant by evolutionary trace method. Results indicate that the, evolutionary trace method like conventional phylogenetic analysis, could resolve phylogenetic relationships between plant haemoglobin family. Evolutionary trace analysis has identified candidate functional (trace) residues that uniquely characterize the heme-binding pocket, dimer interface and possible novel functional surfaces. Such residues from specific three-dimensional clusters might be of functional importance in non-symbiotic haemoglobins. These data, together with our improved knowledge of possible functional residues, can be used in future structure-function analysis experiments.

**Keywords:** evolution; haemoglobin; site directed mutagenesis; phylogeny

#### Background:

Haemoglobins are found ubiquitously in eukaryotes and are also present in many bacteria. Although haemoglobin sequences are highly variable (some sequences share only 20% amino acid sequence identity), they have a conserved structure, the globin fold and often involved in reversible binding of gaseous ligands and the ability to bind other cellular molecules. [1] Haemoglobins are widely distributed in plants and by comparing sequences, expression patterns, and ligand-binding properties, it is now evident that there are at least three distinct types of haemoglobins categorized as symbiotic (Lbs), non-symbiotic (nsHbs) and truncated haemoglobins (tHbs). [2] Symbiotic haemoglobins (Lbs) or leghaemoglobins, the most well known of the plant haemoglobins, are found mainly in plants able to achieve symbiosis with nitrogen-fixing bacteria, i.e., in most legume species (symbiosis with *Rhizobium* and *Bradyrhizobium*) and in several species from *Fagales*, *Rosales*, and *Cucurbitales* (symbiosis with *Frankia*). These symbiotic haemoglobins are expressed at high concentrations in nodules. In these symbiosis-specific structures, bacteria fix nitrogen by nitrogenase enzyme, which requires substantial amounts of energy in the form of ATP produced by bacterial respiration. However, oxygen, which is necessary for respiration, readily inhibits the activity of nitrogenase. Leghaemoglobin owing to its extremely fast O<sub>2</sub> association rate and rather slow O<sub>2</sub> dissociation rate facilitates oxygen diffusion to the symbionts and contributes to the maintenance of an oxygen-free environment around the enzyme. [3]

Though the isolation of haemoglobin from *Parasponia andersonii*, a non leguminous plant, led to the hypothesis that presence of haemoglobin may extend beyond legumes more concrete evidence came after the discovery of barley haemoglobin and the subsequent demonstration that haemoglobin was present in a number of other cereals, such as maize, wheat rye and triticale. [4] To date non-symbiotic haemoglobin genes are identified from both nitrogen and non-nitrogen fixing dicot and from monocot species. Non-symbiotic haemoglobins differ markedly in gene homology from the symbiotic haemoglobins. Current studies indicate that most plants have two or more different non-symbiotic haemoglobin genes that are individually regulated. The function of non-symbiotic haemoglobins in plants is the subject of much current research. These non-symbiotic haemoglobins are expressed at low levels and have expression patterns different to the symbiotic haemoglobins. Based on the predicted oxygen-binding kinetics and probable concentration of the non-symbiotic haemoglobins Appleby et al., (1988) first suggested that the one possible role of the non-symbiotic Hbs could be to sense oxygen levels. [5] Further work indicates that they are expressed differentially, with patterns that are complex and vary from tissue to tissue as well as temporally and in response to different stressors leading to the several speculative functions for non-symbiotic haemoglobins. Some of the suggestions indicate their roles in oxygen storage and transport, detoxification of nitric oxide and/or other reactive oxygen species, and sensing of oxygen or other small heme ligands. Current research shows that, plant non-symbiotic

haemoglobins are involved in nitric oxide signaling. Plants have a general response to many environmental stresses including pathogen attack, wounding and adverse environmental conditions. In response to pathogens, plants also produce an oxidative burst composed of superoxide, hydrogen peroxide and other reactive oxygen species whose production involves nitric oxide. In addition, the expression of plant non-symbiotic haemoglobins is also stimulated by somatic embryogenesis, germination and sucrose, each of which turns on the plant disease resistance pathway, generating nitric oxide and other reactive oxygen species. These proteins cannot facilitate oxygen transport because their affinities are too high and their dissociation kinetics are too slow. Further it is now evident that unlike the symbiotic haemoglobins which show reversible oxygen binding, the non-symbiotic proteins display highly stable oxygen binding and are unlikely to act as oxygen carriers. Hence the physiological function of non-symbiotic haemoglobins in plants must involve something other than traditional oxygen storage and transport. So the exact physiological role of non-symbiotic haemoglobins remains unclear in plants. [2]

Although the details of leghaemoglobin structure and function have been extensively studied, the more recently discovered non-symbiotic plant haemoglobins are still a mystery. The most unusual feature of non-symbiotic haemoglobins is hexacoordination of the heme group. This differs markedly from 'traditional' pentacoordinate haemoglobins, with their open binding site for exogenous ligands. Instead, the heme group in non-symbiotic haemoglobins is coordinated by two histidines similar to the bovine cytochrome b5. The molecular details of this binding reaction are currently one of the principal questions about hexacoordination and haemoglobins. Although such hexacoordination was first observed in the plant non-symbiotic haemoglobins, it has now also been identified in haemoglobins from the photosynthetic microorganisms *Synechocystis* and *Chlamydomonas* that are members of the 'truncated hemoglobin' family. This family contains both pentacoordinate and hexacoordinate haemoglobins that are phylogenetically and structurally different from other haemoglobins and are common to many bacteria and protists. Truncated Hbs (tHbs) are short versions of the globin fold. A putative member of truncated haemoglobins is recently identified in *Arabidopsis*. [6] However, unlike bovine cytochrome b5, which is unreactive towards oxygen and other gaseous ligands, non-symbiotic haemoglobins bind exogenous ligands rapidly and with high affinity. Because hexacoordinate haemoglobins are up regulated by similar conditions in both plants and animals, there might be a common function for these proteins that involves nitric oxide, reactive oxygen species and hypoxia. Owing to their diverse spatio-temporal expression patterns, the biggest challenge now is to decipher the molecular function of these proteins. [7]

Several attempts have been made to decipher the exact function of haemoglobins in plants. Nakajima et al.; used computational methods like ADM (average distance map)

method to predict the folding kinetics of selected plant haemoglobins. [8] In order to identify the function of various amino acid residues, site directed mutagenesis experiments were also performed on conserved Phe40 (B10) which is close to His61 (E7) from rice Hb1. [9] However, the impact of a large number of single or cumulative mutations on the function of non-symbiotic haemoglobins remains to be tested. Since mutating all residues randomly would be an incredibly time and resource consuming process, ways of narrowing down the choice of possible mutational targets must be sought. In this paper I discuss the use of the evolutionary trace (ET) method, to identify potential targets for mutagenesis in plant haemoglobins, with the aim of finding specificity determinants for Lbs and nsHbs from plants. Evolutionary Trace, developed by Olivier Lichtarge, exploits the fact that residues important to the structure or function of a protein strongly tend to be conserved across species. [10] Briefly, The method uses a multiple sequence alignment of a protein family to generate a 'trace' by comparing the consensus sequences for groups of proteins which originate from a common node in phylogenetic tree and are characterized by a common evolutionary time cut-off (ETC) and classifying each residue as one of three types: absolutely conserved, class-specific and neutral. Here, 'class-specific' denotes residues occupying a strictly conserved location in the sequence alignment, but differing in the nature of their conservation between various subgroups. The trace residues identified can then be mapped on to known protein structures, allowing us to identify clusters of important aminoacids and to distinguish between buried and exposed residues. Depending on the ETC value for which a trace is generated, it is possible to maximize the specificity of the analysis over its sensitivity and vice versa. Thus ET analysis allows for a wide range of 'functional resolution'.

### Methodology:

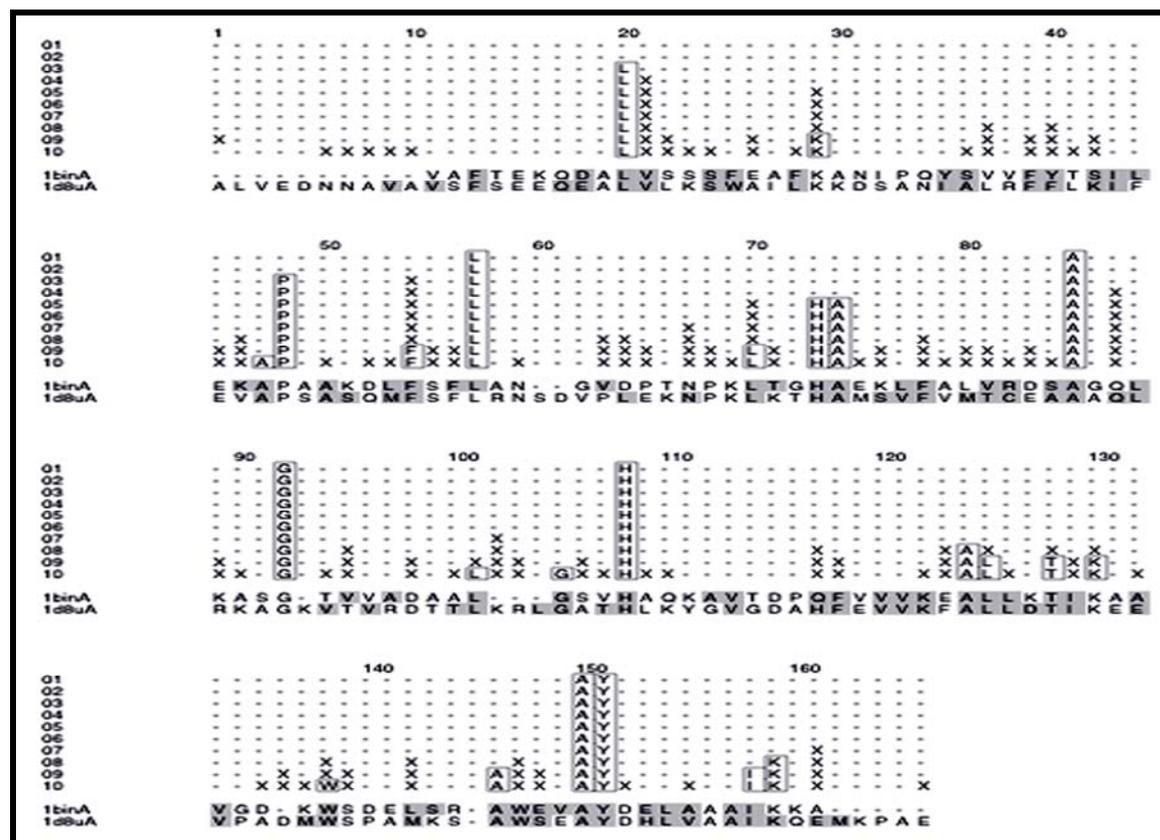
Homologous sequences were obtained with BLASTp [11] using the SWISS PROT [12] protein data base and aligned with ClustalW. [13] After removal of sequences with gaps and outliers of the sequence similarity tree, I obtained an alignment for 74 sequences of the plant haemoglobin family was obtained. Several protein weight matrices such as Blosum 30, Blosum 45, Blosum 62, Gonnet, PAM250 were used to generate the sequence similarity index. However as there were no significant differences observed in conserved amino acid residues with respect to heme binding residues, ET analysis was carried out using BLOSUM62 to allow for substitution of similar amino acids. The aligned sequences and the 1D8U, 1BIN coordinates were submitted to the Cambridge University ET server [14] with default parameters. Briefly the server will then construct a phylogenetic tree based on a phylip distance matrix computed from all the aligned sequences. The sequences on different branches of the phylogram will then be grouped into different evolutionary classes depending on percent similarity. Further to generate these classes, the phylogram will be divided by evolutionary time cutoff lines, into ten evenly distributed partitions P01 to P10, in the order of increasing divergence. In the phylogram, all the sequences that



bryophyte haemoglobins. Node 2 contains all the other plant haemoglobins. In partition P03, plant haemoglobins are differentiated in to symbiotic and non-symbiotic Hbs. With in the non- symbiotic haemoglobin class, class IIb truncated haemoglobins are distinguished from class classI non-symbiotic haemoglobins in partition P05. Partition P07 further differentiates the non-symbiotic haemoglobins from monocots are separated from those of dicots. Thus the ET classes correctly identify all known genetic clusters obtained from conventional phylogenetic analysis. [18] Such clusters lie in different partitions because, unlike conventional phylogentic analysis that determines the clusters by comparing all the sequences together by using arbitrary cutoff values of a sequence similarity index, the ET classes are determined separately for each partition by comparing only those sequences that belong to the nodes originating in that partition. This results in well-defined, non-overlapping ET classes unlike the conventional histogram based cluster analysis that may show overlapping regions in the histogram. Besides many new classes containing sequence similarities that may not be obvious in conventional phylogenetic analysis are created in various ET partitions.

In order to study the structural differences between the known symbiotic and non-symbiotic plant haemoglobins i have first super imposed the crystal structures of rice and soya bean haemoglobins. Known crystal structures for the plant haemoglobin family have relative overall root mean square deviation of  $< 1.1 \text{ \AA}^0$  indicating the conservation of the overall structure with characteristic globin fold. However, Analysis of mapped traces for partitions P01-P10 reveals clusters of potentially important residues appearing on the 3D structures of symbiotic and non-symbiotic plant haemoglobins (Fig.3 A and B). The amino acids are colored according to the relative importance based on the estimated evolutionary pressure they experience. Yellow indicates the highly conserved amino acid residues ( $>95\%$ ) where as white indicates least conserved residues ( $<5\%$ ). The main components of these functional clusters are visible as early as partition P01 and the number of trace residues increase slowly with ETC, thus making it difficult to define cluster boundaries accurately (Fig. 2). Visual inspection of the mapped traces nevertheless led to choosing P08 as the partition displaying the highest ratio of functional resolution over random signal. This is a subjective choice and partitions with both lower and higher ETC values should also be considered, in order to extract the maximum amount of functional information.

### Trace residues and their position on the crystal structures

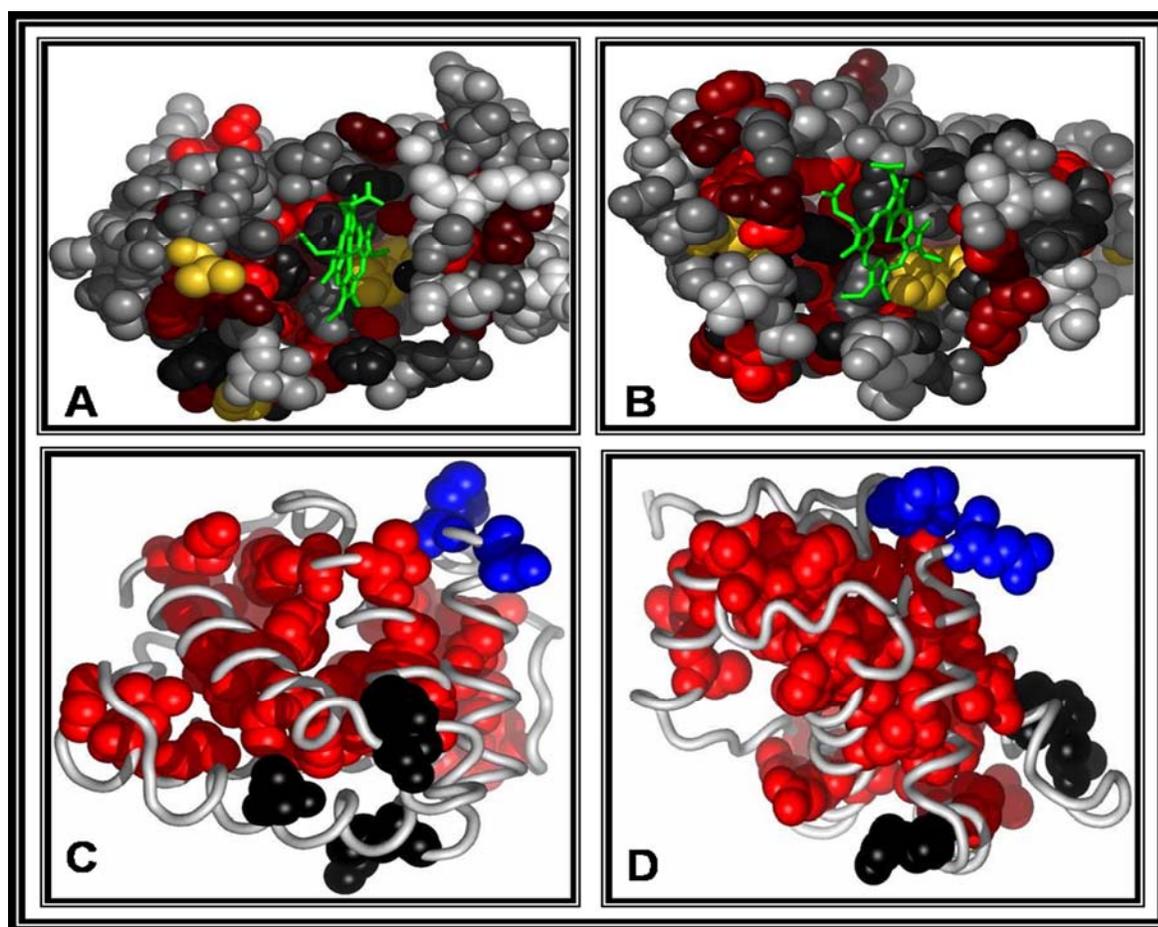


**Figure 2:** Traces for partitions P01-P10, aligned with the amino acid sequences of 1BINA and 1D8U. Conserved residues are surrounded by boxes, class-specific residues are denoted by an X, solvent-accessible side chains are shaded

### Class specific trace residues in plant haemoglobins

Analysis of class-specific residues in the plant haemoglobins is a convenient and objective way to identify features that distinguish the symbiotic and non-symbiotic haemoglobins. Class specificity is determined by sequence conservation within defined subgroups of a protein family. The most striking result of ET analysis is the designation of the catalytic histidine and nearby conserved residues as class specific. Evolutionary trace analysis, could identify the previously know histidine i.e H108 that facilitate the heme iron co-ordination in both symbiotic (pentacoordinate) and non-symbiotic (hexacoordinate) haemoglobins and H73 from non-symbiotic haemoglobin located on E7 and F8 helix positions. Aside from the histidine, many of the class-specific residues identified by the ET surround the positions of the invariant/nearly invariant residues that are not described earlier. These residues include F54, F118, Y150, L153 and I157 in non-symbiotic haemoglobins as conserved in non-symbiotic haemoglobins. In symbiotic haemoglobins at the interface with the heme-iron S45 is conserved which is not conserved in non-symbiotic

haemoglobins. Another significant cluster of class specific residues in plant haemoglobin family is located in the dimer interface. In general plant haemoglobins form homo-dimers in their active form. Evolutionary trace analysis has identified conserved residues that are at the dimer interface. In rice non-symbiotic haemoglobins these residues include E45, A47, P48 and V120. Where as in the symbiotic haemoglobins an additional trace residue W121 is identified at the dimer interface. In addition to the conserved trace residues at the heme-iron binding pocket and the dimer interface, current analysis has identified possible novel functional surfaces (Fig.3 C and D). Table.1 lists all the trace residues and know substitutions at these possible novel functional surfaces in non-symbiotic haemoglobins. The information provided in the 'substitutions' column refers to the other amino acids seen at the same position in the alignment. These amino acid types may be interchangeable at that position in that protein having no effect on the protein function. Hence these residues should be avoided in site directed mutagenesis experiments.



**Figure 3:** Trace residues mapped on to the structure of A. rice non-symbiotic haemoglobin 1D8UA and B. soya bean symbiotic haemoglobin 1BINA. C and D represent novel functional surfaces (shaded in Blue) in symbiotic and non-symbiotic haemoglobins respectively

Residues	Type	Known substitutions (%)
36	A	S(59), G(33), T(2), A(4).
40	F	Y(44), F(55).
55	S	S(93), P(4), T(1).
56	F	F(93), V(1), Y(5).
67	N	S(21), N(77), H(1).
68	P	P(95), T(4).
70	L	L(94), I(1), V(2), A(1).
101	L	L(95), I(2).
158	K	K(93), M(1), I(2).
160	E	A(43), E(51), G(2).

**Table1:** Trace residues that might be of functional signatures in non-symbiotic haemoglobins and known substitutions

### Conclusion:

Evolutionary trace analysis of plant haemoglobin protein sequences has allowed the identification of functional class-specific trace residues that might indicate the functional diversity. Unlike the conventional phylogenetic analysis, evolutionary trace method seeks to identify local class specific patterns of sequence conservation and global patterns of variation that intrinsically represent the functional or structural importance. This method clearly identifies clusters of functionally important residues, outlining the large and complex binding pocket of non-symbiotic haemoglobin thoroughly while making it clear that there are no other areas on the protein structure that are essential for its function. Conserved functional class specific amino acids and the known substitutions reported in the present study could potentially be utilized in future mutational experiments of the plant non-symbiotic haemoglobins.

### References:

- [01] J. B. Wittenberg & B. A. Wittenberg, *Annu Rev Biophys Biophys Chem.*, 19:217 (1990) [PMID: 2194476]
- [02] C. Dordas, *et al.*, *Ann Bot.*, 91:173 (2003) [PMID: 12509338]
- [03] J. B. Wittenberg, *et al.*, *Ann N Y Acad Sci.*, 244:28 (1975) [PMID: 1056166]

- [04] E. R. Taylor, *et al.*, *Plant Mol Biol.*, 24:853 (1994) [PMID: 8204823]
- [05] D. Bogusz, *Nature*, 331:178 (1988) [PMID: 2448639]
- [06] R. A. Watts, *et al.*, *Proc Natl Acad Sci.*, 98:10119 (2001) [PMID: 11526234]
- [07] S. Kundu, *et al.*, *Trends Plant Sci.*, 8:387 (2003) [PMID: 12927972]
- [08] S. Nakajima, *et al.*, *Proteins*, 61:500 (2005) [PMID: 16184600]
- [09] B. J. Smagghe, *et al.*, *Biochemistry*, 45:9735 (2006) [PMID: 16893175]
- [10] O. Lichtarge, *et al.*, *J Mol Biol.*, 257:342 (1996) [PMID: 8609628]
- [11] <http://www.ncbi.nlm.nih.gov/blast/>
- [12] <http://www.expasy.org/>
- [13] J. D. Thompson, *et al.*, *Nucleic Acids Res.*, 22:4673 (1994) [PMID: 7984417]
- [14] <http://www-cryst.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html>
- [15] <http://www.rcsb.org/pdb/>
- [16] <http://www.pymol.org>
- [17] I. Mihalek, *et al.*, *Bioinformatics*, 22:1656 (2006) [PMID:16644792]
- [18] A. B. Heckmann, *et al.*, *Plant Mol Biol.*, 61:769 (2006) [PMID:16897491]

Edited by M. K. Sakharkar

Citation: Reddy, *Bioinformatics* 1(9): 370-375 (2007)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.