JPP JOURNAL OF PROTEINS AND PROTEOMICS
4(3), December 2013, pp. 231-248

Research Article

AN EVALUATION OF POTENTIAL INTRINSICALLY DISORDERED AND AMYLOIDOGENIC REGIONS IN HEMOGLOBINS

Suneeta Basireddy[#], Sheetal Uppal[#], Amit Kumar Singh and Suman Kundu*

Department of Biochemistry, University of Delhi South Campus, Benito Juarez Road, New Delhi 110021, India

Abstract: Discovery of intrinsically disordered proteins (IDPs) has challenged the traditional protein structure-function paradigm. IDPs lack stable tertiary and/or secondary structures and are highly abundant in proteome. Functional diversity of IDPs complements the functions of ordered proteins. Intriguingly, some IDPs share a propensity to form insoluble, ordered aggregates called amyloid and it has been proposed that potential amyloidogenic and disordered regions in polypeptide chains influence the conformation of proteins leading to the formation of such fibers and plaques which are hallmarks of neurodegenerative and many other protein conformational diseases. Recently, several studies have shown the ubiquitous presence of hemoglobin in life forms suggesting a wide variety of biological functions. However, there has been no report for intrinsically disordered hemoglobins or disordered regions in hemoglobins considering the burgeoning number of new hemoglobins, although myoglobin has been shown to have flexible regions. *In silico* tools were used to explore the possibility of presence of disordered regions in different hemoglobins from various organisms. Additionally, attempt was also made to identify amyloidogenic regions, if any, in hemoglobins. Such analysis revealed that several globins potentially display disorder in CD loop region, known to be critical for hemoglobin function and heme stability. *In silico* analysis also predicted the presence of putative amyloidogenic regions in newly discovered hemoglobins. These findings question the structural or functional relevance, demanding detailed investigation, of such disorders in hemoglobins, their implications in pathogenesis through amyloid formation and the role of heme binding in preventing formation of disordered regions or amyloids.

Keywords: Intrinsic disorder in hemoglobins; amyloid; CD loop disorder; amyloidogenic B and G helix.

Introduction

Proteins perform a plethora of functions within living organisms. This diversity in function has always been attributed to the large variety of three-dimensional tertiary structures they can fold into from a linear chain of amino acids (Lodish *et al.*, 2000). This paradigm of widely acclaimed structure-function relationship prevailed for a long time until recently when researchers found that the irregularly structured proteins are the causative agents in various diseases thus leading to increased interest in understanding the role of unstructured proteins (Dobson, 1999; Lemieux and Spohr, 1994; Shakhnovich *et al.*, 2003; Siskova, 2013; Uversky, 2003). Due to their inherent nonstructured segments, which are essential for functions involved in various cellular processes, they were termed as intrinsically disordered proteins (IDPs) (Dunker *et al.*, 2001; Maity and Maiti, 2012; Romero *et al.*, 2001; Williams *et al.*, 2001). It is now known that ~20% proteins in prokaryotes and >50% proteins in eukaryotes are IDPs (Dunker *et al.*, 2001; Dunker and Obradovic, 2001; Sun *et al.*, 2011; Uversky, 2013a) while the number keeps expanding signifying the importance of these class of proteins to life.

Corresponding Author: **Suman Kundu**
E-mail: suman.kundu@south.du.ac.in

Received: November 9, 2013
Accepted: December 24, 2013
Published: December 31, 2013

IDPs lack well defined tertiary and/or secondary structure and fail to form a globular fold to form a nicely folded 3D structure under physiological conditions. IDPs and similar regions in a globular protein, with otherwise stable structure, may rapidly interconvert between several energetically comparable structures (Huang and Grzesiek, 2010; Marsh and Forman-Kay, 2009; Ozenne *et al.*, 2012; Wang *et al.*, 2013). Upon binding to structured proteins or other interacting partners, IDPs undergo disorder to order transition (Espinoza-Fonseca, 2009; Fuxreiter and Tompa, 2012; Maity and Maiti, 2012; Moritsugu *et al.*, 2012; Uversky, 2013b; Wang *et al.*, 2013). This dynamic nature of IDPs provides intrinsic plasticity thus allowing them to recognize different biological targets while still being specific for a particular function.

While the knowledge of crucial diverse functions of IDPs is important, regulation of abundance of IDPs is very critical to prevent any non-specific and non-functional interactions that can affect normal cellular processes (Babu *et al.*, 2011). There is emerging evidence that IDPs are often associated with amyloidosis (Linding *et al.*, 2004; Uversky, 2009; Zerovnik, 2011). Some of the well known IDPs in human alpha–synuclein and tau are implicated in major protein conformational diseases (Shastry, 2003; Skrabana *et al.*, 2006; Uversky, 2009). In IDPs, hydrophobic sequences are exposed to the solvent in absence of other interacting protein partners and thus increase the amyloidogenicity of the corresponding protein (Douglas and Douglas, 2010; Uversky, 2008).

Despite advances in our knowledge of protein folding, we know and understand much less about protein aggregation and misfolding, two related processes that often have undesirable consequences both *in vivo* and *in vitro.* Today, there are more than 40 serious human diseases which are associated with aggregation of proteins (Chiti and Dobson, 2006; Dobson, 1999; Eisenberg and Jucker, 2012; Luheshi and Dobson, 2009). The *in vivo* formation of fibrillar proteinaceous deposits with a motif known as cross beta sheet is termed as amyloid (Fitzpatrick et al., 2013; Squires *et al.*, 2006). Initially, it was believed that protein aggregation diseases are exclusive only to central nervous system but recent studies showed that proteins in peripheral tissues are also involved in protein aggregation diseases (Koo *et al.*, 1999; Stefani and Dobson, 2003). Such diseases include type-2 diabetes, inherited cataracts, some forms of atherosclerosis, hemodialysis-related disorders, and short-chain amyloidosis, among many others (Crabbe, 1998; Dobson, 2006). During the past few years, investigations have also shown that there have been an increasing number of proteins with no link to protein deposition diseases that under certain *in vitro* conditions, form fibrillar aggregates that have the morphological, structural, and tinctorial properties that allow them to be classified as amyloid fibrils (Bauer *et al.*, 1995; Pedersen *et al.*, 2006; Saiki *et al.*, 2005; Serpell *et al.*, 2000; Sunde and Blake, 1997; Sunde *et al.*, 1997). This finding has led to the idea that the ability to form amyloid structure is an inherent or generic property of polypeptide chains. It is possible that continued investigation might link such proteins to aggregation diseases. In the least, they provide valuable model systems for understanding the mechanism of amyloid formation.

With regard to such significance of IDPs and amyloidogenic proteins in general, a class of proteins which has been relatively less investigated is the hemoglobin family, although the paradigm myoglobin has been shown to have flexible F-helix (Fontana *et al.*, 2004). Details for intrinsically disordered segments and amyloidosis propensity of hemoglobins have not been reported per se, except for myoglobin (Griko Iu *et al.*, 1988; Kim *et al.*, 2005), probably because these proteins are known to be compact and highly structured consisting of a typical all alpha-helical globin fold wrapped around a heme iron center. Over the past few years, bulging population of new hemoglobins have been discovered in all forms of life with varying sequence, fold and structure, active pocket architecture, plasticity, cellular localization and function (Kumar *et al.*, 2013; Milani *et al.*, 2001; Pesce *et al.*, 2000; Vinogradov *et al.*, 2006; Vinogradov *et al.*, 2013). Investigations have shown that in addition to classical transporting, storing O_2 and facilitating its diffusion, these hemoglobins exhibit functions that range from control of nitric oxide (NO) levels in

microorganisms to use of NO to control the level of O_2 in nematodes, binding and transport of sulfide in endosymbiont-harboring species, protection against sulfide, scavenging of O_2 in symbiotic leguminous plants, O_2 sensing in bacteria and archaebacteria, dehaloperoxidase activity useful in detoxification of chlorinated materials, electron transport, signaling through G-proteins, tumor suppressor, neuroprotection, etc. (Cheung *et al.*, 2004; Fordel *et al.*, 2007; Greenberg *et al.*, 2008; Hebelstrup *et al.*, 2007; Shivapurkar *et al.*, 2008; Vinogradov *et al.*, 2005; Wakasugi *et al.*, 2003; Weber and Vinogradov, 2001; Yu *et al.*, 2012). Their ubiquitous nature supports such multiple and essential physiological roles. Since presence of disordered segments in proteins help them assume multiple roles, it was imperative that several members of the new hemoglobin family be explored for the presence of such regions, and their potential association to the phenomenon of amyloidosis. Such investigation assumes significance in light of the fact that apomyoglobin has been used in the recent past to explore the mechanism and principles of amyloid fibril formation (Fandrich *et al.*, 2003; Fandrich *et al.*, 2006; Picotti *et al.*, 2007; Sirangelo *et al.*, 2004; Vilasi *et al.*, 2006). The issue of generality of such process for hemoglobins across species and with multiple functions prompted us to investigate the disordered and amyloidogenic regions across globins from different organisms. In the current investigation, we employed an *in silico* approach to identify disordered and amyloidogenic regions in different hemoglobins (Figure 1).

Methods

Hemoglobin sequences were retrieved from NCBI (*http://www.ncbi.nlm.nih.gov/*). Relevant crystal structures of several hemoglobins were obtained (18 out of 40) from PDB (*http://www.rcsb.org/pdb/home/home.do*) and viewed in PyMOL (*http://www.pymol.org/*), a molecular visualization and analysis program (DeLano, 2002).

Prediction for disordered regions: DISOPRED2 Prediction of Protein Disorder Server (*http://bioinf.cs.ucl.ac.uk/psipred/*) was used for such prediction (Ward *et al.*, 2004).

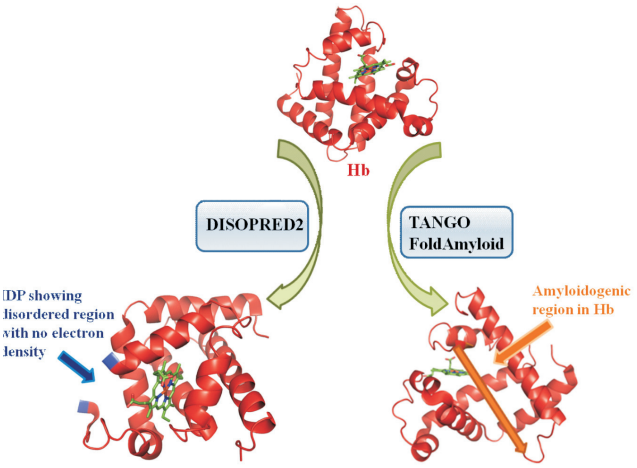

Figure 1: **Online servers used for predicting the intrinsic disorder and amyloidogenic regions in hemoglobins.**

Prediction for amyloidogenic regions: Tango [Waltz] (*http://waltz.switchlab.org/*) and FoldAmyloid (*http://bioinfo.protres.ru/fold-amyloid/oga.cgi*) servers (Fernandez-Escamilla *et al.*, 2004; Garbuzynskiy *et al.*, 2010; Linding *et al.*, 2004; Rousseau *et al.*, 2006) were employed for prediction of segments that can result in amyloid formation.

For hemoglobins whose crystal structures were not available, sequence alignment with hemoglobin sequences of known structures was performed using Multialin (*http://multalin.toulouse.inra.fr/multalin/*) software (Corpet, 1988) to analyze the nature of segments that produce disorder and amyloid sensitivity.

Secondary structure predictions for segments of hemoglobins were performed using PSIPRED (*http://bioinf.cs.ucl.ac.uk/psipred/*) (McGuffin *et al.*, 2000).

Results and Discussion

Disordered Regions in Hemoglobins are Mostly Housed in the CD Loop

Hemoglobins are heme containing all alpha-helical proteins with a high order of compactness (Dickerson and Geis, 1983). Hence global disorder is not expected in these proteins at least in their holo forms. Apoglobins are more likely to show global disorder, if any. One would thus expect local disorders, if any, in apohemoglobins, as has been shown previously for F-helix of

apomyoglobin (Fontana *et al.*, 2004). Hence, in the current investigation *in silico* tools were used to predict both disorder and propensity of amyloidosis in hemoglobins (Figure 1), as a prelude to future experimental efforts. Forty (40) hemoglobins across species were used for such analysis (Table 1).

The analysis reveals the presence of disordered regions of varying length in almost all the different globins inspected (Table 1). With the exception of three globins – eye globin (chicken), *Paramecium* globin and *Synechocystis* globin – it appears that all hemoglobins have potential local segments of disorder. What separates these globins from others is a matter of speculation and needs further investigation. It appears that disorder is independent of heme coordination (penta- or hexa- coordination), species or class of hemoglobins.

Intriguingly, majority of the hemoglobins (29 out of 37) displayed disorder in the CD loop and this result seems to be a general feature for globins. Some of them also possess disordered regions in N- and C- terminal, as can be expected for a protein. In search of a valid experimental proof of these findings of specific disordered regions, it was realized that missing electron density in protein crystal structures usually arise from failure to solve the phase problem, from crystal defects, or even from unintentional proteolytic processing during protein purification (Dunker *et al.*, 2001). The common reason for missing electron density is that the unobserved atom, side chain, residue, or region fails to scatter X-rays coherently due to variation in position from one protein to the next, i.e., the unobserved atoms are disordered (Dunker *et al.*, 2001). If the hemoglobins above are predicted to have disordered regions in CD loop, N- or C-terminal, then crystal structures of these globins might exhibit absence of electron density in these regions. With such a rationale crystal structures of some of these globins were closely inspected (Figure 2). Indeed we observed that the crystal structures of most of the globins showed no electron density in these regions (Figure 2). The potential disorder predicted in hemoglobins could thus be significant and demands further investigation.

Extensive studies on globins indicate that CD1 Phe is conserved across hemoglobins in different organisms (Couture *et al.*, 1999; Pesce *et al.*, 2000). *In silico* analysis here concludes that CD region being unstructured is very much flexible. In hemoglobins, CD loop undergoes significant rearrangements which mediate side chain displacement around heme upon ligand binding (Emerson *et al.*, 1988). There are several reports which suggest that CD region is important in controlling the ligand affinity but that effect of rearrangement is specific to each protein (Lopez and Kollman, 1993; Whitaker, 1996). Plant hemoglobins (*Arabidopsis thaliana* class 1, 2 and 3 non-symbiotic Hbs, etc) and insect hemoglobins (*Drosophila melanogaster* Hb1) also showed disorder in CD region suggesting that this flexibility is conserved and has functional significance (de Sanctis *et al.*, 2005; Mukhi *et al.*, 2013; Trevaskis *et al.*, 1997). The worth of these regions in hemoglobins could be recognized and appreciated by deleting these flexible segments or subjecting them to mutational analysis to see their effect on globin structure and function. These local regions of disorder could potentially mediate interactions with other proteins leading to multiple functions that hemoglobins are now thought to be capable of.

Hemoglobins Constitute Amyloidogenic Regions in Key Helices

The potential presence of local segments of intrinsic disorder in hemoglobins encouraged us to investigate amyloidogenic regions in them as well, since many IDPs are also known to form amyloid fibrils (Linding *et al.*, 2004; Uversky, 2009; Zerovnik, 2011). Moreover, previously it has been reported that apomyoglobin forms amyloid fibrils by association of unfolded polypeptide segments under partially denaturing conditions (Fandrich *et al.*, 2003). However, no other hemoglobin has been extensively investigated along similar lines and the generality of this phenomenon is not known. If it is well established that all hemoglobins form amyloid fibrils under appropriate physiochemical environment, the case for their pathogenic significance can acquire momentum. It is also interesting to know which segments are involved in the formation of amyloid fibrils in hemoglobins and their relation

Table 1
Prediction of Disordered Regions in Various Hemoglobins

Sl. No	Hemoglobin	Organism Type	Heme Coordination	Disordered Regions (amino acid positions and loops and helices)	PDB ID Referred
1	*Arabidopsis thaliana* class 1 non-symbiotic hemoglobin	Plant (Dicot)	Hexa	Major:1-9#, 55-67 CD loop to E helix Minor:159, 160#	3ZHW
2	*Arabidopsis thaliana* class 2 non-symbiotic hemoglobin	Plant (Dicot)	Hexa	Major:1-4*, 53-64 CD loop to E helix, 155-158* region after H helix	
3	*Casuarina glauca* non-symbiotic hemoglobin II class1	Plant (Dicot)	NA	Major:1-8#, 55-67 CD loop to E helix Minor:158-160* region after H helix	
4	*Casuarina glauca* hemoglobin class2	Plant (Dicot)	NA	Major:49-61 CD loop to E helix Minor:1*, 152 E helix	
5	*Chlamydomonas eugametos* hemoglobin (Truncated)	Green algae	Hexa	Major:1-8, 16-44** Minor:162-164# C terminal region	1DLY
6	*Cichorium* hemoglobin I class2	Plant (Dicot)	Hexa	Major: 49-61 CD loop to E helix, 151-154 E helix 155-161# Minor: 1#	
7	*Cichorium* hemoglobin II class2	Plant (Dicot)	Hexa	Major: 49-61 CD loop to E helix, 152, 153 E helix, 154-165# Minor: 1#	
8	Human Cytoglobin	Mammal	Hexa	Major:1-18* region before A helix, 174-190* region after H helix	1UMO
9	*Drosophila melanogaster* hemoglobin 1	Insect	Hexa	Major:50-57 CD loop-E helix Minor:152 C terminal loop	2G3H
10	Eye Globin [*Gallus gallus* (chicken)]	Bird	Penta	Nil	
11	*Glycine max* lba class2	Plant (Dicot)	Penta	Major:50-59 CD loop to D helix Minor:1, 2* region before A helix	1FSL
12	*Glycine max* lbc class2	Plant (Dicot)	Penta	Major:49-59 CD loop to D helix Minor:1, 2* region before A helix	
13	*Hordeum vulgare* non-symbiotic hemoglobin class1	Plant (Monocot)	Hexa	Major:1-4* region before A helix, 56-67 C helix, CD loop Minor:160-162* region after H helix	2OIF
14	Human hemoglobin Alpha	Mammal	Penta	Major:52 CD loop Minor:1* region before A helix	1GZX
15	Human hemoglobin Beta	Mammal	Penta	Minor:55 D helix	1GZX
16	Human Myoglobin	Mammal	Penta	Major:54-59 D helix, DE loop Minor:154* region after H helix	3RGK
17	*Lotus japonicus* leghemoglobin	Plant (Dicot)	NA	Major: 48-59 CD loop to E helix Minor: 1#	
18	*Lupinus luteus* LbI class2	Plant (Dicot)	Penta	Major:51-61CD loop, D helix Minor:1, 2* region before A helix, 152-154* region after H helix	

19	*Lupinus luteus* LbII class2	Plant (Dicot)	Penta	Major:48-58 CD loop, D helix Minor:150, 151 region after H helix	2GDM
20	*Marchantia polymorpha* hemoglobin class0	Liverwort	NA	Major: 29-42 CD loop to E helix, 139-145#, 155-161# Minor: 1#	
21	*Medicago sativa* lba class2	Plant (Dicot)	NA	Major: 47-58 CD loop to E helix Minor: 1#	
22	*Medicago sativa* Lbc class2	Plant (Dicot)	NA	Major: 49-60 CD loop to E helix Minor: 1#	
23	Human Neuroglobin	Mammal	Hexa	Minor:150, 151 C terminal loop	1OJ6
24	*Oryza sativa* hemoglobinI class1	Plant (Monocot)	Hexa	Major:1-14 N terminal loop, 60-72 CD loop-E helix, 164-166 C terminal loop	1D8U
25	*Oryza sativa* hemoglobinII class1	Plant (Monocot)	Hexa	Major:1-18 N terminal loop, 63-74 CD loop-E helix, 167-169 C terminal loop	
26	*Paramecium caudatum* hemoglobin	Protozoa (Ciliate)	NA	Nil	1DLW
27	*Parasponia andersonii* hemoglobin class1	Plant	Penta	Major:1-6# region before A helix, 56-68 CD helix to E helix, 159-162# region after H helix Minor:8, 9 # region before A helix	3QQR
28	*Physcomitrella patens* hemoglobin class0	Plant (Bryophyte)	NA	Major: 1-25#, 71-81 CD loop to E helix, 175 E helix 176-180#	
29	*Psophocarpus tetragonolobus* hemoglobin	Plant (Dicot)	NA	Major:49-60 CD loop to E helix Minor:1, 2* region before A helix	
30	*Rattus norvegicus* Neuroglobin	Mammal	Hexa	Minor:151* region after H helix	1Q1F
33	*Sesbania rostrata* leghemoglobin	Plant (Dicot)	NA	Major: 50-60 CD loop to E helix Minor: 1#	
34	Sperm Whale Myoglobin	Mammal	Penta	Major:53-60 D helix Minor:1*	1MBN
35	*Synechocystis* hemoglobin	Cyanobacteria	Hexa	Nil	1RTX
36	*Vicia sativa* leghemoglobin	Plant (Dicot)	Penta	Minor:1* region before A helix	
37	*Vigna unguiculata* hemoglobin	Plant (Dicot)	NA	Major: 48-61 B helix to CD loop Minor: 1, 2#	
38	*Xenopus tropicalis* globin X	Amphibian	NA	Major: 3-28# region before A helix, 78-85 CD helix, D helix, 176-179 H helix, 180-189* region after H helix Minor: 1#	
39	*Zea mays* hemoglobinI classx	Plant (Monocot)	Hexa	Major:1-11# region before a helix, 60-63#**64-70 region between C helix and E helix, 162-165 C terminal loop	2R50
40	*Zea mays* hemoglobinII classx	Plant (Monocot)	Hexa	Major:50-61 #** region between C helix and E helix Minor:1#	

*** Region with no secondary structural data assigned**
#Region with no data available
****Complete protein structure except this region was available. NA – not available**
Hexa-Hexacoordinated
Penta-Pentacoordinated

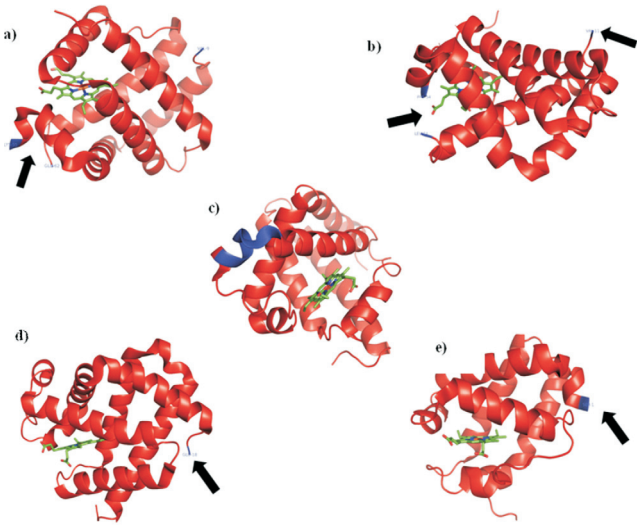

Figure 2: **Crystal structures of hemoglobins showing regions (bold arrows) that lack electron density, which also coincide with predicted regions of local disorder. a)** ***Parasponia andersonii*** **Hb (PDB ID- 3QQR (Kakar** ***et al.*****, 2011)), showing absence of initial 8 residues and residues from 56-61. b)** ***Zea mays*** **HbI (PDB ID- 2R50 (Smagghe** ***et al.*****, In Press)), showing absence of region between C helix and E helix. c) Sperm whale myoglobin (PDB ID- 1MBN (Watson, 1969)), showing flexible regions between residues 53 and 60. d) Human cytoglobin (PDB ID- 1UMO (de Sanctis** ***et al.*****, 2004)), showing absence of initial 17 residues. e)** ***Chlamydomonas eugametos*** **Hb (PDB ID- 1DLY (Pesce** ***et al.*****, 2000), showing truncation in A helix.**

to those that result in disorder. For apomyoglobin, it has been reported that the peptide fragment corresponding to the G- helix or N terminal region may be playing a role for the alpha to beta transition such that the globular protein goes into amyloid fibrils (Fandrich *et al.*, 2003; Picotti *et al.*, 2007). The scientific curiosity as to whether other helices in hemoglobins are also involved motivated the current investigation.

Available online servers were used for predicting amyloidogenic regions across the globin proteins. The robustness of such servers are still questionable since they often predict multiple regions in a protein and as such two *in silico* tools were used and only the results (regions) that were common to both were considered significant. As a positive control, myoglobin was used and indeed A- and G- helices were predicted to be amyloidogenic *in silico* as was observed *in vitro*, thus instilling confidence in further analysis. Sixty (60) globins were subjected to analysis and the results are summarized in Table 2.

Table 2 clearly indicates that presence of amyloidogenic regions is common in hemoglobins. Only ten globins showed absence of amyloidogenic region but that too only with one of the servers (Tango). The other server (FoldAmyloid), however, predicted such regions in all the 60 globins investigated. It is important to note, however, that FoldAmyloid server predicts multiple amyloidogenic regions in most hemoglobins, which seems somewhat of overprediction. Hence emphasis was more on the regions that were predicted in common by both the servers.

Interestingly, it was found that more than half of the hemoglobins (35 out of 60) showed amyloidogenic region in their B-Helix. About 20% of the hemoglobins (12 out of 60) displayed disorder in G-helix resembling myoglobin. Only 2 Hbs, both from plants (*Hordeum vulgare* non-symbiotic Hb class1 and *Oryza sativa* HbII class1), showed amyloidogenic region in A-helix like myoglobin. Only 2 Hbs exhibited amyloidogenic region in C- and D-helices, one in each. Segments in CD region were found to be amyloidogenic in 9 Hbs, although majority of the globins showed local disorder in this region (Table 1). E-, F- and H- helices showed amyloidogenic region in 5, 3 and 7 globins, respectively. One Hb showed amyloid forming region in FG region (*Thermobifida fusca* Hb) as well. Table 3 summarizes the regions that are amyloidogenic showing the number of such globins for each of the segments.

It is evident from Table 3 and Table 1, that some globins have common regions where both local disorder and amyloidogenicity is predicted, namely the CD loop. However, such commonality is predicted in less than $1/6^{th}$ of the globins investigated, indicating that the disorder in CD loop is important for function as discussed above, and the probability of this region being amyloidogenic as well maybe low. Somewhat higher probability for amyloid formation (second topmost hit in Table 3) was displayed by the G-helix. This segment was reported to be amyloidogenic for apomyoglobin *in vitro* (Fandrich *et al.*, 2003; Picotti *et al.*, 2007) and seems more likely. The most significant segment for amyloid formation in hemoglobins seems to lie

Table 2
Predicted Amyloidogenic Regions in Hemoglobins

Sl. No	Hemoglobin	FoldAmyloid Prediction (aa location)	Tango Prediction (aa location)	Common Amyloidogenic region	PDB ID Referred
1.	*Agrobacterium tumefaciens* Hb	19 —— 28 52 —— 56 75 —— 83 92 —— 99 111 —— 117 120 —— 125	NRP	NCR	2XYK
2.	*Arabidopsis thaliana* class 1 non-symbiotic Hb	33 —— 42 50 —— 54 71 —— 78 118 —— 123 146 —— 151	35-42	35-42 (B-Helix and CD Region)	3ZHW
3.	*Arabidopsis thaliana* class 2 non-symbiotic Hb	29 —— 35 47 —— 51 68 —— 73 98 —— 102 143 —— 149	35-40 45-49	47-49 (CD loop)	2GDM# (Lupinus luteus LbII class2)
4.	*Bacillus subtilis* Hb	22 —— 27 51 —— 56 71 —— 75 77 —— 81 88 —— 92 107 —— 113 118 —— 122	NRP	NCR	1UX8
5.	*Campylobacter jejuni* Hb	15 —— 20 52 —— 59 80 —— 89 104 —— 108 111 —— 115 118 —— 123	2-12 14-19 82-91	5-19 (B-Helix) 82-89 (G-Helix)	2IG3
6.	*Casuarina glauca* non-symbiotic HbII	33 —— 42 48 —— 52 71 —— 77 118 —— 123	35-43	35-42 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
7.	*Casuarina glauca* Hb class 2	27 —— 35 42 —— 46 65 —— 70 95 —— 99	29-36 42-46 64-72	29-35 (B-Helix) 42-46 (CD Region) 65-70 (E-Helix)	2GDM# (Lupinus luteus LbII class2)
8.	*Cerebratulus lacteus* Hb (nemertean worm)	1 —— 6 9 —— 14 51 —— 55	NRP	NCR	1KR7
9.	*Chironomus thummi thummi* erythrocruorin	22 —— 28 63 —— 68 126 —— 135	21-29 103-108	22-28 (B-Helix)	1ECD
10.	*Chlamydomonas eugametos* Hb	1 —— 9 11 —— 15 73 —— 77 89 —— 95 121 —— 125	71-79 87-95	73-77 (D-Helix) 89-95 (E-Helix)	1DLY
11.	*Cichorium* I class2 HbI	26 —— 35 42 —— 46 65 —— 72	24-37 42-48	26-35 (B-Helix) 42-46 (CD Region)	2GDM# (Lupinus luteus LbII class2)
12.	*Cichorium* I classx HbII	26 —— 32 65 —— 70 143 —— 147	24-37 142-146	26-32 (B-Helix) 143-146 (H-Helix)	2GDM# (Lupinus luteus LbII class2)

13.	*Drosophila* Glob1	26 —— 34 63 —— 68 86 —— 91 110 —— 115 138 —— 144	28-36 137-143	28-34 (B-Helix) 138-143 (H-helix)	2G3H
14.	*Drosophila* Glob2	63 —— 67 80 —— 85 91 —— 95 110 —— 115 127 —— 132 150 —— 155 158 —— 163 180 —— 185	77-85 89-93 181-190	80-85 (F-Helix) 91-93 (F-Helix) 181-185**	1UMO# (Human Cytoglobin)
15.	*Eptatretus burgeri* Hb ((hagfish)	34 —— 44 72 —— 77 111 —— 115 119 —— 123 135 —— 144	71-86 109-115 117-126 134-146	72-77 (E-Helix) 111-115 (G-Helix) 119-123 (G-Helix) 135-144 (H-Helix)	1IT2
16.	Eye globin	16 —— 20 29 —— 35 42 —— 46 85 —— 89 108 —— 118	69-74 107-116	108-116 (G-Helix)	1UMO# (Human Cytoglobin)
17.	*Gasterophilus intestinalis* (botfly) Hb	29 —— 33 40 —— 46 112 —— 120 133 —— 138 140 —— 146	140-151	140-146 (H-Helix)	2C0K
18.	*Geobacillus stearothermophilus* Hb	4 —— 8 21 —— 26 50 —— 55 70 —— 80 87 —— 94 108 —— 121	NRP	NCR	2BKM
19.	*Glycine max* Lba class2	27 —— 33 45 —— 49 65 —— 70	27-35	27-33 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
20.	*Glycine max* non-symbiotic Hbclass1	33 —— 42 48 —— 54 71 —— 77 118 —— 123	35-43	35-42 (B and C-Helix)	2GDM# (Lupinus luteus LbII class2)
21.	*Hordeum vulgare* (Barley) non-symbiotic Hb class1	17 —— 24 33 —— 43 49 —— 55 72 —— 78 115 —— 124	6-12 20-24 36-44	20-24 (A-Helix) 36-43 (B-Helix)	2OIF
22.	Human alpha chain	31 —— 37 42 —— 49 99 —— 110	NRP	NCR	——————
23.	Human beta chain	13 —— 17 30 —— 38 40 —— 45 104 —— 118	127-132	NCR	——————
24.	Human Cytoglobin	28 —— 36 43 —— 52 58 —— 62 121 —— 126 128 —— 134 156 —— 161	41-54 122-134	43-52 (B-Helix) 122-126 (G-Helix)	1UMO
25.	Human Myoglobin	8 —— 14 29 —— 35 102 —— 107 109 —— 115 137 —— 141	27-32 104-116	29-32 (B-Helix) 104-115 (G-Helix)	3RGK

26.	Human Neuroglobin	25 —— 34 39 —— 48 66 —— 72 113 —— 119	24-35	25-34 (B-Helix)	1OJ6
27.	*Lotus japonicus* leghemoglobin	27 —— 35	25-37	25-35 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
28.	*Lupinus luteus* LbI class2	28 —— 36	133-140	NCR	————
29.	*Lupinus luteus* LbII class2	25 —— 33 66 —— 72	130-137	NCR	
30.	*Marchantia polymorpha* Hb	7 —— 16 22 —— 26 45 —— 50 128 —— 132	12-17 22-26 127-132	12-17 (B-Helix) 22-26 (CD Region) 128-132 (H-Helix)	2GDM# (Lupinus luteus LbII class2)
31.	*Medicago sativa* Lba class2	26 —— 33 89 —— 93	24-34	26-33 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
32.	*Medicago sativa* Lbc class2	26 —— 34 91 —— 95 102 —— 106	25-36 42-48	26-34 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
33.	*Mycobacterium tuberculosis* HbN	1 —— 6 30 —— 37 43 —— 47 60 —— 64 86 —— 93	41-48 57-63	43-47 (CD Region) 60-63 (E-Helix)	1IDR
34.	*Mycobacterium tuberculosis* HbO	19 —— 25 48 —— 56 76 —— 83 87 —— 94 113 —— 118	NRP	NCR	1NGK
35.	*Oryza sativa* HbI class1	21 —— 28 36 —— 46 53 —— 59 76 —— 82 119 —— 128 152 —— 157	40-46	40-46 (B-Helix)	1D8U
36.	*Oryza sativa* HbII class1	24 —— 31 40 —— 49 56 —— 62 79 —— 85 122 —— 131	13-19 27-31 43-49	27-31 (A-Helix) 43-49 (B-Helix)	1D8U*
37.	*Paramecium caudatum* Hb	1 —— 5 18 —— 23 31 —— 35 48 —— 52	17-25	18-23 (B-Helix)	1DLW
38.	*Parasponia andersonii* symbiotic Hb class2	17 —— 24 34 —— 43 73 —— 78 102 —— 106 115 —— 124 150 —— 155	33-44 49-55 142-147	34-43 (B-Helix)	3QQR
39.	*Peromyscus maniculatus* (Deer mouse)Hb	30 —— 35 41 —— 48 98 —— 109	NRP	NCR	4H2L
40.	*Petromyzon marinus* (Lamprey) Hb	36 —— 43 70 —— 74 114 —— 118 138—— 147	26-30 37-45 47-53 74-79	37-43 (B-Helix)	1F5P

41.	*Phacoides pectinatus* Hb	28 —— 34	NRP	NCR	1FLP
42.	*Physcomitrella patens* Hb	49 —— 57 65 —— 69 85 —— 91 136 —— 140	53-58 134-142	53-57 (B-Helix) 136-140 (G-Helix)	2GDM# (Lupinus luteus LbII class2)
43.	*Psophocarpus tetragonolobus* Hb	27 —— 33 45 —— 49	27-35 130-138	27-33 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
44.	*Rattus Norvegicus* Neuroglobin	8 —— 12 25 —— 34 39 —— 44 66 —— 72 113 —— 119	24-35	25-34 (B-Helix)	1Q1F
45.	*Riftia pachyptila* C1 Hb	2 —— 6 30 —— 38 103 —— 114 120—— 124	34-39 103-108 130-136	34-38 (B-Helix)	1YHU
46.	*Scapharca* tetrameric Hb	35 —— 44 50 —— 54 72 —— 80	76-81 98-103 110-114	NCR	4HRT
47.	*Sesbania rostrata* leghemoglobin	27 —— 34 103 —— 107	25-35	27-34 (B-Helix)	2GDM# (Lupinus luteus LbII class2)
48.	Sperm Whale Myoglobin	8 —— 17 29 —— 35 69 —— 73 104 —— 109 114 —— 120 139 —— 143	106-116	104-109 (G-Helix) 114-116 (G-Helix)	1MBN
49.	*Synechocystis* Hb	32 —— 36 49 —— 55	49-56	49-55 (E-Helix)	1RTX
50.	*Tetrahymena pyriformis* Hb	22 —— 27 35 —— 39 52 —— 58 80 —— 85	NRP	NCR	3AQ6
51.	*Thermobifida fusca* (Actinobacterium)	1 —— 5 14 —— 22 46 —— 54 69 —— 78 85 —— 92 106—— 120	106-123	106-120 (F-Helix, FG Region, G-Helix)	2BMM
52.	*Thunnus thynnus Hb* (bluefin tuna)	13 —— 17 31 —— 36 42 —— 46 89 —— 93 100—— 113 129—— 138	NRP	NCR	1V4U
53.	*Tokunagayusurika akamusi* Hb	30 —— 35 67 —— 76 113—— 121 134—— 143	29-37 62-75 112-117 133-141	30-35 (B-Helix) 67-75 (E-Helix) 113-117 (G-Helix) 134-141 (H-Helix)	1X3K
54.	*Trema tomentosa* non symbiotic Hb	1 —— 6 23 —— 30 40 —— 49 78 —— 84 108—— 112 121—— 130	42-50 55-61	42-49 (B-Helix, CD Region)	3QQQ
55.	*Trematomus bernacchii* Hb	11 —— 16 30 —— 36 41 —— 45 99 —— 113	32-37 106-112	32-36 (B-Helix) 106-112 (G-Helix)	1S5X

56.	*Vicia sativa* leghemoglobin	25 —— 32	NRP	NCR	————
57.	*Vigna unguiculata* (Cowpea) Hb	27 —— 34 45 —— 49	27-35 44-50	27-34 (B-Helix) 45-49 (CD Loop)	2GDM# (Lupinus luteus LbII class2)
58.	Globin X [*Xenopus tropicalis*]	1 —— 6 8 —— 12 36 —— 47 53 —— 62 69 —— 75 94 —— 99 132 —— 136 140 —— 144 162 —— 172	53-61 68-73 163-169	53-61 (B-Helix) 69-73 (CD Loop) 163-169 (H-Helix)	1UMO# (Human Cytoglobin)
59.	*Zea mays* HbI classx	20 —— 27 36 —— 46 52 —— 58 75 —— 81 108 —— 112 122 —— 127 136 —— 14	42-47 146-152	42-46 (B-Helix)	2R50
60.	*Zea mays* HbII classx	10 —— 14 27 —— 36 42 —— 48 66 —— 72 115—— 119 126—— 130	32-37	32-36 (B-Helix)	2R50*

NRP- No region predicted
NCR-No common region
****Region with no sequence similarity with the template structure**
#1UMO (Human Cytoglobin) and 2GDM (*Lupinus luteus* LbII class2) used as template for the hemoglobins with no available PDB structure.
aa – amino acid

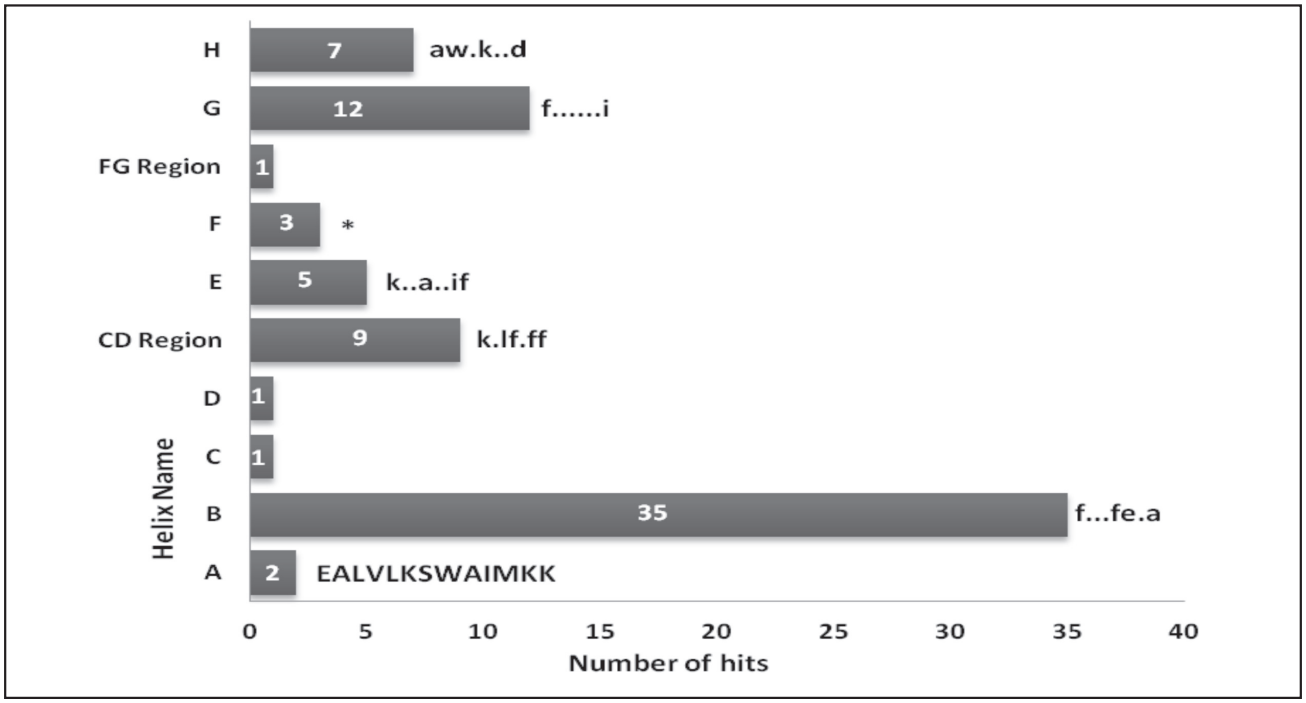

Figure 3: **Number of hits of disordered sequences obtained for each helix in 40 Hbs analysed. For each helix all the hits obtained were aligned and the consensus sequence obtained is written in bold.**
***No sequence similarity obtained**

in the B-helix (Table 3). It is worthwhile to investigate some of these globins to verify this finding *in vitro*, especially since of late the B-helix gained increasing importance in hemoglobins in their functional and structural relevance (Kundu *et al.*, 2004; Kundu and Hargrove, 2003; Kundu *et al.*, 2002; Smagghe *et al.*, 2006). That E- and F-helices were also predicted to be amyloidogenic in a few globins is the most surprising result since these helices are responsible for housing the heme and plays the most significant role in ligand binding in hemoglobins (Scott *et al.*, 2001; Vallone *et al.*, 2004). It is possible that they play a role in amyloidogenesis, if any, only in apoglobins and not hologlobins where they are engaged in other functions, and offers scope for interesting future investigation. In some cases the H-helix was also predicted and seems likely source of helix to sheet transition for amyloid formation. In summary, atleast five of possible eight helices in hemoglobins showed significant potential to undergo amyloid transition albeit with different probability and appears that A-, C-, and D- helices are less likely, with B-helix being the most likely candidate followed by G-helix and CD loop.

It can be argued that if the individual segments in hemoglobins have the tendency to undergo amyloid transition, then these sequence stretches when taken out of the context of the whole polypeptide chain and their structural and functional relevance, may display some propensity to form β-sheets. This was verified for the individual segments in many types of hemoglobin by predicting their secondary structure *in silico*. Majority of these regions were found to have propensity to form β-sheets, a requisite in amyloid structures (data not shown).

For experimental verification of the findings here, it will be necessary to perform mutational analysis in some of these hemoglobins, especially in the regions that are predicted to be amyloidogenic. Hence we aligned sequences of B-helices of all the 35 Hbs that showed amyloidogenic region in this stretch of the polypeptide chain. Interestingly, a specific Phe was found to be conserved in most of the Hbs in the amino acid region of B-helix that was most amyloidogenic (Figure 3). Likewise, sequences of other helices were also aligned and analyzed (Figure 3). In G-helix, Phe and Ile were found to be conserved. Phe was again found to be conserved in the CD loop region. Our findings thus indicate that Phe, a hydrophobic residue, might be playing some sort of key role in amyloidogenesis which could be either in initiation or promotion of amyloid formation (Maity and Maiti, 2012). In H-helix, a Trp was found to be conserved which is present upstream to the predicted amyloid segment but is not a part of it. This suggests that the hydrophobic residues present in the helices which are not a part of the predicted amyloid nuclues may also play a role in amyloid formation. A- and F- helices were found to be least amyloidogenic.

In silico mutational analysis of the potential hydrophobic residue(s) in predicted amyloidogenic region showed a decrease and in some cases even complete suppression of amyloidogenesis using the same predictive tools as used for the wild type proteins above. For example, Y52A mutation in *Synechocystis* Hb results in predictive scores of almost nil for the corresponding helix to be amyloidogenic. Such hypothesis can be confirmed by *in vitro* site-directed mutational studies, helping the cause of investigation of the mechanism of amyloid formation in hemoglobins in particular and proteins in general.

Conclusion

The investigation indicates the presence of intrinsically disordered and amyloidogenic regions in hemoglobins and they are generic in nature. The disorder regions are mostly present in CD loop region and may be important for their function in ligand binding and protein-protein interaction. Other regions that display such disorder could also be located at the protein terminals. The *in silico* observations find resonance in the fact that experimental crystal structures of many hemoglobins lack electron density in their CD loop regions or polypeptide terminus indicating disorder that prevents well defined- X-ray diffraction pattern. Only a small fraction of the local disorders were also predicted to be amyloidogenic in nature. It appears that the regions that are most likely to result in amyloid formation in hemoglobins are located in B-helix

Table 3
Summary of the Number of Globins with Amyloidogenic Region in each Helix

Helix	Number of hits	Globins
A	2	*Hordeum vulgare* non-symbiotic Hb class1, *Oryza sativa* HbII class1
B	35	*Arabidopsis thaliana* class 1 non-symbiotic Hb, *Campylobacter jejuni* Hb, *Casuarina glauca* non-symbiotic HbII, *Casuarina glauca* Hb class 2, *Chironomus thummi thummi* erythrocruorin, *Cichorium* I class2 HbI, *Cichorium* I classx HbII, Human Cytoglobin, *Drosophila* Glob1, *Glycine max* Lba class2, *Glycine max* non-symbiotic Hbclass1, *Hordeum vulgare* non-symbiotic Hb class1, Human Myoglobin, *Lotus japonicus* leghemoglobin, *Marchantia polymorpha* Hb, *Medicago sativa* Lba class2, *Medicago sativa* Lbc class2, Human Ngb, *Oryza sativa* HbI class1, *Oryza sativa* HbII class1, *Paramecium caudatum* Hb, *Parasponia andersonii* symbiotic Hb class2, *Petromyzon marinus*, *Physcomitrella patens* Hb, *Psophocarpus tetragonolobus* Hb, *Rattus Norvegicus* Neuroglobin, *Riftia pachyptila* C1 Hb, *Sesbania rostrata* leghemoglobin, *Tokunagayusurika akamusi* Hb, *Trema tomentosa* non symbiotic Hb, *Trematomus bernacchii* Hb, *Vigna unguiculata* (Cowpea) Hb, Globin X [*Xenopus tropicalis*], *Zea mays* HbI classx, *Zea mays* HbII classx
C	1	*Glycine max* non-symbiotic Hbclass1
D	1	*Chlamydomonas eugametos* Hb
CD region	9	*Arabidopsis thaliana* class 1 non-symbiotic Hb, *Arabidopsis thaliana* class 2 non-symbiotic Hb, *Casuarina glauca* Hb class 2, *Cichorium* I class2 HbI, *Marchantia polymorpha* Hb, *Mycobacterium tuberculosis* HbN, *Trema tomentosa* non symbiotic Hb, *Vigna unguiculata* (Cowpea) Hb, Globin X [*Xenopus tropicalis*]
E	5	*Casuarina glauca* Hb class 2, *Chlamydomonas eugametos* Hb, *Mycobacterium tuberculosis* HbN, *Synechocystis* Hb, *Tokunagayusurika akamusi* Hb
F	3	*Drosophila* Glob2, *Thermobifida fusca* (Actinobacterium)
FG Region	1	*Thermobifida fusca* (Actinobacterium)
G	12	*Campylobacter jejuni* Hb, Human Cytoglobin, *Eptatretus burgeri* Hb ((hagfish), Eye globin, Human Myoglobin, *Physcomitrella patens* Hb, Sperm Whale Myoglobin, *Thermobifida fusca* (Actinobacterium), *Tokunagayusurika akamusi* Hb, *Trematomus bernacchii* Hb
H	7	*Cichorium* classx HbII, DmeGlob1, *Eptatretus burgeri* Hb ((hagfish), *Gasterophilus intestinalis* (botfly) Hb, *Marchantia polymorpha* Hb, *Tokunagayusurika akamusi* Hb, Globin X [*Xenopus tropicalis*]

followed by those in G-helix. While there is experimental evidence of the latter from studies of apomyoglobin, those for the former are wanting. The *in silico* analyses help formulate several testable hypotheses that can be verified experimentally to understand whether most hemoglobins can form amyloids and if they do so, what is the mechanism of such phenomenon.

One disadvantage of the analysis presented here is the inability to clearly distinguish whether disorder or amyloid formation will result from holo- or apo- forms of globins, though it is intuitive that hologlobins are less likely to do so due to their compact structures. The predictions are indeed based on hemoglobin sequences, without factoring in heme binding, and would thus hold true for apoglobins, if at all. Ofcourse, the disorders in CD loop regions observed in crystal structures are only in holo forms. In apo-forms those disorders would tend to be higher and heme binding would probably neutralize many such local disorders. As for amyloid formation, reports exist only for apomyoglobin and hologlobins would probably resist such structural transitions unless they lose heme under certain physiological conditions. Nonethelesss, disorder and amyloidogenesis indeed present exciting opportunities in hemoglobin research.

Abbreviations

Hb, Hemoglobin; IDPs, Intrinsically disordered proteins; ID, Intrinsic disorder; NO, Nitric Oxide; O_2, Oxygen.

Acknowledgement

SB and SU acknowledge UGC and CSIR, Government of India for research fellowships. Central Instrumentation Facility at University of Delhi South Campus is appreciated for bioinformatics facility. Financial assistance from UGC (SAP programme), DBT (COE programme) and Delhi University (R&D grant) are duly acknowledged.

References

Babu, M.M., van der Lee, R., de Groot, N.S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol *21*, 432-440.

Bauer, H.H., Aebi, U., Haner, M., Hermann, R., Muller, M., and Merkle, H.P. (1995). Architecture and polymorphism of fibrillar supramolecular assemblies produced by in vitro aggregation of human calcitonin. J Struct Biol *115*, 1-15.

Cheung, B.H., Arellano-Carbajal, F., Rybicki, I., and de Bono, M. (2004). Soluble guanylate cyclases act in neurons exposed to the body fluid to promote C. elegans aggregation behavior. Curr Biol *14*, 1105-1111.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem *75*, 333-366.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res *16*, 10881-10890.

Couture, M., Yeh, S.R., Wittenberg, B.A., Wittenberg, J.B., Ouellet, Y., Rousseau, D.L., and Guertin, M. (1999). A cooperative oxygen-binding hemoglobin from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *96*, 11223-11228.

Crabbe, M.J. (1998). Cataract as a conformational disease—the Maillard reaction, alpha-crystallin and chemotherapy. Cell Mol Biol (Noisy-le-grand) *44*, 1047-1050.

de Sanctis, D., Dewilde, S., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T., and Bolognesi, M. (2004). Crystal structure of cytoglobin: the fourth globin type discovered in man displays heme hexa-coordination. J Mol Biol *336*, 917-927.

de Sanctis, D., Dewilde, S., Vonrhein, C., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T., Ponassi, M., Nardini, M., *et al.* (2005). Bishistidyl heme hexacoordination, a key structural property in Drosophila melanogaster hemoglobin. J Biol Chem *280*, 27222-27229.

DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.

Dickerson, R.E., and Geis, I. (1983). Hemoglobin: Structure, Function, Evolution and Pathology. Menlo Park, CA: The Benjamin/Cummings Publishing Co, Inc.

Dobson, C.M. (1999). Protein misfolding, evolution and disease. Trends Biochem Sci *24*, 329-332.

Dobson, C.M. (2006). Protein aggregation and its consequences for human disease. Protein Pept Lett *13*, 219-227.

Douglas, P.M., and Douglas, M.C. (2010). Interplay Between Protein Homeostasis Networks in Protein Aggregation and Proteotoxicity. Biopolymers *93*, 229-236.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., *et al.* (2001). Intrinsically disordered protein. J Mol Graph Model *19*, 26-59.

Dunker, A.K., and Obradovic, Z. (2001). The protein trinity—linking function and disorder. Nat Biotechnol *19*, 805-806.

Eisenberg, D., and Jucker, M. (2012). The amyloid state of proteins in human diseases. Cell *148*, 1188-1203.

Emerson, S.D., Lecomte, J.T., and LaMar, G.N. (1988). Proton NMR resonance assignment and dynamic analysis of phenylalanine CD1 in a low-spin ferric complex of sperm whale myoglobin. J Am Chem Soc *110*, 4176-4182.

Espinoza-Fonseca, L.M. (2009). Reconciling binding mechanisms of intrinsically disordered proteins. Biochem Biophys Res Commun *382*, 479-482.

Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C.M., and Diekmann, S. (2003). Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. Proc Natl Acad Sci U S A *100*, 15463-15468.

Fandrich, M., Zandomeneghi, G., Krebs, M.R., Kittler, M., Buder, K., Rossner, A., Heinemann, S.H., Dobson, C.M., and Diekmann, S. (2006). Apomyoglobin reveals a random-nucleation mechanism in amyloid protofibril formation. Acta Histochem *108*, 215-219.

Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol *22*, 1302-1306.

Fitzpatrick, A.W., Debelouchina, G.T., Bayro, M.J., Clare, D.K., Caporini, M.A., Bajaj, V.S., Jaroniec, C.P., Wang, L., Ladizhansky, V., Muller, S.A., *et al.* (2013). Atomic structure and hierarchical assembly of a cross-beta amyloid fibril. Proc Natl Acad Sci U S A *110*, 5468-5473.

Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structutre by limited proteolysis. Acta Biochimica Polonica *51*, 299-321.

Fordel, E., Thijs, L., Moens, L., and Dewilde, S. (2007). Neuroglobin and cytoglobin expression in mice. Evidence for a correlation with reactive oxygen species scavenging. FEBS J *274*, 1312-1317.

Fuxreiter, M., and Tompa, P. (2012). Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol *725*, 1-14.

Garbuzynskiy, S.O., Lobanov, M.Y., and Galzitskaya, O.V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics *26*, 326-332.

Greenberg, D.A., Jin, K., and Khan, A.A. (2008). Neuroglobin: an endogenous neuroprotectant. Curr Opin Pharmacol *8*, 20-24.

Griko Iu, V., Privalov, P.L., Ven'iaminov, S., and Kutyshenko, V.P. (1988). [Thermodynamic study of the structure of apomyoglobin]. Biofizika *33*, 18-26.

Hebelstrup, K.H., Igamberdiev, A.U., and Hill, R.D. (2007). Metabolic effects of hemoglobin gene expression in plants. Gene *398*, 86-93.

Huang, J.R., and Grzesiek, S. (2010). Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. J Am Chem Soc *132*, 694-705.

Kakar, S., Sturms, R., Tiffany, A., Nix, J.C., DiSpirito, A.A., and Hargrove, M.S. (2011). Crystal structures of Parasponia and Trema hemoglobins: differential heme coordination is linked to quaternary structure. Biochemistry *50*, 4273-4280.

Kim, Y.J., Kim, Y.A., Park, N., Son, H.S., Kim, K.S., and Hahn, J.H. (2005). Structural characterization of the molten globule state of apomyoglobin by limited proteolysis and HPLC-mass spectrometry. Biochemistry *44*, 7490-7496.

Koo, E.H., Lansbury, P.T., Jr., and Kelly, J.W. (1999). Amyloid diseases: abnormal protein aggregation in neurodegeneration. Proc Natl Acad Sci U S A *96*, 9989-9990.

Kumar, A., Nag, M., and Basak, S. (2013). Truncated or 2/2 Hemoglobins : A new class of globins with novel structure and function. Journal of Proteins and Proteomics *4*, 45-64.

Kundu, S., Blouin, G.C., Premer, S.A., Sarath, G., Olson, J.S., and Hargrove, M.S. (2004). Tyrosine B10 inhibits stabilization of bound carbon monoxide and oxygen in soybean leghemoglobin. Biochemistry *43*, 6241-6252.

Kundu, S., and Hargrove, M.S. (2003). Distal heme pocket regulation of ligand binding and stability in soybean leghemoglobin. Proteins *50*, 239-248.

Kundu, S., Snyder, B., Das, K., Chowdhury, P., Park, J., Petrich, J.W., and Hargrove, M.S. (2002). The leghemoglobin proximal heme pocket directs oxygen dissociation and stabilizes bound heme. Proteins *46*, 268-277.

Lemieux, U.R., and Spohr, U. (1994). How Emil Fischer was led to the lock and key concept for enzyme specificity. Adv Carbohydrate Chem Biochem *10*, 1-20.

Linding, R., Schymkowitz, J., Rousseau, F., Diella, F., and Serrano, L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol *342*, 345-353.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, D. (2000). "Hierarchical Structure of Proteins" Molecular Cell Biology, 4th edition, New York: W.H. Freeman.

Lopez, M.A., and Kollman, P.A. (1993). Application of molecular dynamics and free energy perturbation methods to metalloporphyrin-ligand systems II: CO and dioxygen binding to myoglobin. Protein Sci *2*, 1975-1986.

Luheshi, L.M., and Dobson, C.M. (2009). Bridging the gap: from protein misfolding to protein misfolding diseases. FEBS Lett *583*, 2581-2586.

Maity, M., and Maiti, N.C. (2012). Sequence Composition of Binding Sites in Natively unfolded Human Proteins. Journal of Proteins and Proteomics *3*, 119-127.

Marsh, J.A., and Forman-Kay, J.D. (2009). Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. J Mol Biol *391*, 359-374.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

Milani, M., Pesce, A., Ouellet, Y., Ascenzi, P., Guertin, M., and Bolognesi, M. (2001). Mycobacterium tuberculosis hemoglobin N displays a protein tunnel suited for O2 diffusion to the heme. EMBO J *20*, 3902-3909.

Moritsugu, K., Terada, T., and Kidera, A. (2012). Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. J Am Chem Soc *134*, 7094-7101.

Mukhi, N., Dhindwal, S., Uppal, S., Kumar, P., Kaur, J., and Kundu, S. (2013). X-ray crystallographic structural characteristics of Arabidopsis hemoglobin I and their functional implications. Biochim Biophys Acta *1834*, 1944-1956.

Ozenne, V., Schneider, R., Yao, M., Huang, J.R., Salmon, L., Zweckstetter, M., Jensen, M.R., and Blackledge, M. (2012). Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. J Am Chem Soc *134*, 15138-15148.

Pedersen, J.S., Dikov, D., Flink, J.L., Hjuler, H.A., Christiansen, G., and Otzen, D.E. (2006). The changing face of glucagon fibrillation: structural polymorphism and conformational imprinting. J Mol Biol *355*, 501-523.

Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L., and Bolognesi, M. (2000). A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family. EMBO J *19*, 2424-2434.

Picotti, P., De Franceschi, G., Frare, E., Spolaore, B., Zambonin, M., Chiti, F., de Laureto, P.P., and Fontana, A. (2007). Amyloid fibril formation and disaggregation of fragment 1-29 of apomyoglobin: insights into the effect of pH on protein fibrillogenesis. J Mol Biol *367*, 1237-1245.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins *42*, 38-48.

Rousseau, F., Schymkowitz, J., and Serrano, L. (2006). Protein aggregation and amyloidosis: confusion of the kinds? Curr Opin Struct Biol *16*, 118-126.

Saiki, M., Honda, S., Kawasaki, K., Zhou, D., Kaito, A., Konakahara, T., and Morii, H. (2005). Higher-order molecular packing in amyloid-like fibrils constructed with linear arrangements of hydrophobic and hydrogen-bonding side-chains. J Mol Biol *348*, 983-998.

Scott, E.E., Gibson, Q.H., and Olson, J.S. (2001). Mapping the pathways for O2 entry into and exit from myoglobin. J Biol Chem *276*, 5177-5188.

Serpell, L.C., Sunde, M., Benson, M.D., Tennent, G.A., Pepys, M.B., and Fraser, P.E. (2000). The protofilament substructure of amyloid fibrils. J Mol Biol *300*, 1033-1039.

Shakhnovich, B.E., Dokholyan, N.V., DeLisi, C., and Shakhnovich, E.I. (2003). Functional fingerprints of folds: evidence for correlated structure-function evolution. J Mol Biol *326*, 1-9.

Shastry, B.S. (2003). Neurodegenerative disorders of protein aggregation. Neurochem Int *43*, 1-7.

Shivapurkar, N., Stastny, V., Okumura, N., Girard, L., Xie, Y., Prinsen, C., Thunnissen, F.B., Wistuba, II, Czerniak, B., Frenkel, E., *et al.* (2008). Cytoglobin, the newest member of the globin family, functions as a tumor suppressor gene. Cancer Res *68*, 7448-7456.

Sirangelo, I., Malmo, C., Iannuzzi, C., Mezzogiorno, A., Bianco, M.R., Papa, M., and Irace, G. (2004). Fibrillogenesis and cytotoxic activity of the amyloid-forming apomyoglobin mutant W7FW14F. J Biol Chem *279*, 13183-13189.

Siskova, Z. (2013). How structure shapes (dys)function: A perspective to understanding brain region-specific degeneration in prion disease. Prion *7*.

Skrabana, R., Sevcik, J., and Novak, M. (2006). Intrinsically disordered proteins in the neurodegenerative processes: formation of tau protein paired helical filaments and their analysis. Cell Mol Neurobiol *26*, 1085-1097.

Smagghe, B.J., Hoy, J.A., and Hargrove, M.S. (In Press). The crystal structure of nonsymbiotic corn hemoglobin 1.

Smagghe, B.J., Kundu, S., Hoy, J.A., Halder, P., Weiland, T.R., Savage, A., Venugopal, A., Goodman, M., Premer, S., and Hargrove, M.S. (2006). Role of phenylalanine B10 in plant nonsymbiotic hemoglobins. Biochemistry *45*, 9735-9745.

Squires, A.M., Devlin, G.L., Gras, S.L., Tickler, A.K., MacPhee, C.E., and Dobson, C.M. (2006). X-ray scattering study of the effect of hydration on the cross-beta structure of amyloid fibrils. J Am Chem Soc *128*, 11738-11739.

Stefani, M., and Dobson, C.M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J Mol Med (Berl) *81*, 678-699.

Sun, X., Xue, B., Jones, W.T., Rikkerink, E., Dunker, A.K., and Uversky, V.N. (2011). A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. Plant Mol Biol *77*, 205-223.

Sunde, M., and Blake, C. (1997). The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Protein Chem *50*, 123-159.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. J Mol Biol *273*, 729-739.

Trevaskis, B., Watts, R.A., Andersson, C.R., Llewellyn, D.J., Hargrove, M.S., Olson, J.S., Dennis, E.S., and Peacock, W.J. (1997). Two hemoglobin genes in Arabidopsis thaliana: the evolutionary origins of leghemoglobins. Proc Natl Acad Sci U S A *94*, 12230-12234.

Uversky, V.N. (2003). Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell Mol Life Sci *60*, 1852-1871.

Uversky, V.N. (2008). Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res *5*, 260-287.

Uversky, V.N. (2009). Intrinsic disorder in proteins associated with neurodegenerative diseases. Front Biosci (Landmark Ed) *14*, 5188-5238.

Uversky, V.N. (2013a). A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci *22*, 693-724.

Uversky, V.N. (2013b). Intrinsic disorder-based protein interactions and their modulators. Curr Pharm Des *19*, 4191-4213.

Vallone, B., Nienhaus, K., Matthes, A., Brunori, M., and Nienhaus, G.U. (2004). The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. Proc Natl Acad Sci U S A *101*, 17351-17356.

Vilasi, S., Dosi, R., Iannuzzi, C., Malmo, C., Parente, A., Irace, G., and Sirangelo, I. (2006). Kinetics of amyloid aggregation of mammal apomyoglobins and correlation with their amino acid sequences. FEBS Lett *580*, 1681-1684.

Vinogradov, S.N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J.R. (2006). A phylogenomic profile of globins. BMC Evol Biol *6*, 31.

Vinogradov, S.N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Guertin, M., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J.R. (2005). Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. Proc Natl Acad Sci U S A *102*, 11385-11389.

Vinogradov, S.N., Tinajero-Trejo, M., Poole, R.K., and Hoogewijs, D. (2013). Bacterial and archaeal globins - a revised perspective. Biochim Biophys Acta *1834*, 1789-1800.

Wakasugi, K., Nakano, T., and Morishima, I. (2003). Oxidized human neuroglobin acts as a heterotrimeric Galpha protein guanine nucleotide dissociation inhibitor. J Biol Chem *278*, 36505-36512.

Wang, Y., Chu, X., Longhi, S., Roche, P., Han, W., Wang, E., and Wang, J. (2013). Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proc Natl Acad Sci U S A *110*, E3743-3752.

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., and Jones, D.T. (2004). The DISOPRED server for the prediction of protein disorder. Bioinformatics *20*, 2138-2139.

Watson, H.C. (1969). The Stereochemistry of the Protein Myoglobin. ProgStereochem *4*, 299.

Weber, R.E., and Vinogradov, S.N. (2001). Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol Rev *81*, 569-628.

Whitaker, T.L. (1996). Residues Controlling the Function and Stability of the CD corner in Myoglobin and Hemoglobin, *Ph.D. Dissertation*. Rice University, Houston, TX.

Williams, R.M., Obradovi, Z., Mathura, V., Braun, W., Garner, E.C., Young, J., Takayama, S., Brown, C.J., and Dunker, A.K. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac Symp Biocomput, 89-100.

Yu, Z., Liu, N., Liu, J., Yang, K., and Wang, X. (2012). Neuroglobin, a Novel Target for Endogenous Neuroprotection against Stroke and Neurodegenerative Disorders. Int J Mol Sci *13*, 6995-7014.

Zerovnik, E. (2011). Oligomerization preceding amyloid fibril formation: a process in common to intrinsically disordered and globular proteins. Network *22*, 154-161.