Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Understanding the roles of intrinsic disorder in subunits of hemoglobin and the disease process of sickle cell anemia

Reis Fitzsimmons[a], Narmin Amin[a], and Vladimir N. Uversky[a,b,c]

[a]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA; [b]USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA; [c]Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

ABSTRACT

One of the common genetic disorders is sickle cell anemia, in which 2 recessive alleles must meet to allow for destruction and alteration in the morphology of red blood cells. This usually leads to loss of proper binding of oxygen to hemoglobin and curved, sickle-shaped erythrocytes. The mutation causing this disease occurs in the 6th codon of the *HBB* gene encoding the hemoglobin subunit β (β-globin), a protein, serving as an integral part of the adult hemoglobin A (HbA), which is a heterotetramer of 2 α chains and 2 β chains that is responsible for binding to the oxygen in the blood. This mutation changes a charged glutamic acid to a hydrophobic valine residue and disrupts the tertiary structure and stability of the hemoglobin molecule. Since in the field of protein intrinsic disorder, charged and polar residues are typically considered as disorder promoting, in opposite to the order-promoting non-polar hydrophobic residues, in this study we attempted to answer a question if intrinsic disorder might have a role in the pathogenesis of sickle cell anemia. To this end, several disorder predictors were utilized to evaluate the presence of intrinsically disordered regions in all subunits of human hemoglobin: α, β, δ, ε, ζ, γ1, and γ2. Then, structural analysis was completed by using the SWISS-MODEL Repository to visualize the outputs of the disorder predictors. Finally, Uniprot STRING and D^2P^2 were used to determine biochemical interactome and protein partners for each hemoglobin subunit along with analyzing their posttranslational modifications. All these properties were used to determine any differences between the 6 different types of subunits of hemoglobin and to correlate the mutation leading to sickle cell anemia with intrinsic disorder propensity.

Introduction

Sickle cell anemia or sickle cell disease/disorder is an autosomal recessive genetic disease, the form of the congenital hemoglobinopathy, that is caused by the "substitution of one amino acid in the hemoglobin molecule."[1] The World Health Organization estimated in 2006 that 5 percent of the world population carries a gene for a hemoglobinopathy. About 300,000 children with this disease are born each year,[2,3] with two-thirds of these births being in Africa.[4] In the UK, it is estimated that there are 12,000–15,000 affected individuals and over 300 infants born with sickle cell disease in the UK each year who are diagnosed as part of the neonatal screening program.[5] In the United States, it is estimated that sickle cell anemia is present in 1 in 500 livebirths among Americans of African descent, 1 in 12 African American have the trait, and approximately 100,000 Americans largely of African descent live with the disease.[6]

This disease is caused by the sickle cell transformation of erythrocytes (red blood cells, RBCs), which can no longer properly bind to oxygen. Low oxygen levels can cause "occlusion of blood vessels, increased viscosity, and inflammation."[1] Sickle cell was the first genetic disorder to be "identified at the molecular level" in 1957.[7] The reason was that it was caused by the substitution of a native glutamic acid to valine in the sixth codon of the human *HBB* gene encoding the hemoglobin subunit β (β-globin), a protein, serving as an integral part of the adult hemoglobin A (HbA), which is a heterotetramer of 2 α chains and 2 β chains that is responsible for binding to the oxygen in the blood. Homozygotes for sickle cell mutation have abnormal hemoglobin which "polymerizes in long

fibers" when red blood cells lose their oxygen supply.[7] This is a major factor that explains how the RBCs transform into sickle-shaped, deformed floppy discs. Although the reason might sound very insignificant at first, the mutation creates radical changes in the structure and function of the RBCs. When the glutamic acid residue is replaced by valine, the position for a charged residue is replaced with a nonpolar residue, which could "cause some disruption of the tertiary structure."[8] Arends et al. mentioned that when oxygen levels were measured in a heterozygote, they tended to be normal, but when the oxygen levels were compared to those of a recessive homozygote, there was decreased affinity for oxygen likely determined by the disruption in the tertiary structure of hemoglobin.[8] The lowering of the oxygen affinity levels, lead to the reshaping of the red blood cells into a new dysfunctional morphology that suspends their activity of carrying oxygen. The glutamic acid might be influential because it is a strategically placed charged residue that can play a role in enforcing the normal structure of the hemoglobin. However, when it has been replaced by valine, the protein becomes more hydrophobic and likely to increase the predisposition of this subunit for self-aggregation.

There is a growing amount of evidence which suggests that many protein regions, and even entire proteins, lack a stable tertiary and/or secondary structure in solution, and instead exist as dynamic ensembles of interconverting structures. Therefore, functional proteins can be grouped into several general classes, such as transmembrane proteins, globular proteins, fibrous proteins, intrinsically disordered proteins (IDPs), and hybrid proteins with intrinsically disordered protein regions (IDPRs). Although these IDPs and IDPRs are biologically active, they still fail to form specific 3D structures, and instead exist as extended dynamically mobile conformational or collapsed ensembles.[9-16] Therefore, unlike ordered proteins, whose 3-D structure is relatively stable and whose Ramachandran angles vary only slightly around their equilibrium positions with occasional cooperative conformational switches, IDPs/IDPRs exist as structural ensembles, either at the secondary or at the tertiary level. Similar to ordered proteins, which are able to correctly fold into relatively rigid biologically active conformations based on their amino acid sequences, the lack of rigid structure in IDPs/IDPRs is also encoded in the specific features of their amino acid sequences. In fact, some of

these proteins were discovered due their unusual amino acid sequence compositions. The absence of regular structure in these proteins has been explained by the specific features of their amino acid sequences, including the presence of numerous uncompensated charged groups (often negative); i.e., a high net charge at neutral pH, which is a result of the extreme pI values in such proteins,[17-19] and a low content of hydrophobic amino acid residues.[17,18] More focused analysis of the IDPs/IDPRs revealed that they are significantly depleted in bulky hydrophobic (Ile, Leu, and Val) and aromatic amino acid residues (Trp, Tyr, and Phe), which would normally form the hydrophobic core of a folded globular protein, and also possess a low amount of Cys and Asn residues. The depletion of disordered protein in Cys is also crucial, as this amino acid residue is known to have a significant contribution to the protein conformation stability via disulfide bond formation or by being involved in the coordination of different prosthetic groups. It has been proposed that these depleted residues, Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn be called order-promoting amino acids. On the other hand, IDPs were shown to be substantially enriched in Ala, as well as in polar, disorder-promoting amino acids: Arg, Gly, Gln, Ser, Glu, and Lys, and also in the hydrophobic, but structure-breaking Pro.[12,20-23]

Curiously, sickle cell anemia is caused by the substitution of a disorder-promoting glutamic acid to an order-promoting valine. This observation defined the objective of this study, which was the better understanding of the roles of intrinsic disorder in normal hemoglobin and in sickle cell anemia-related mutant. Other considered factors were posttranslational modifications and biochemical interactions with other proteins.

Materials and methods

Amino acid sequences of all proteins analyzed in this study (in FASTA format) and some general information related to their structure and function were retrieved from UniProt.[24] Multiple sequence alignment was conducted using the alignment internet tool Clustal Omega (1.2.1).[25]

Intrinsic disorder propensities of target proteins were evaluated using 4 algorithms from the PONDR family, PONDR-FIT and PONDR® VSL2,[26-28] as well as the IUPred web server.[29] For each protein, after

obtaining an average disorder score by each predictor, all predictor-specific average scores were averaged again to generate an average per-protein intrinsic disorder score. Use of consensus for evaluation of intrinsic disorder is motivated by empirical observations that this approach usually increases the predictive performance compared to the use of a single predictor.[30-32] We further characterized the disorder status of query proteins using the MobiDB database (http://mobidb.bio.unipd.it/),[33,34] that generates consensus disorder scores by aggregating the output from 10 predictors, such as 2 versions of IUPred,[29] 2 versions of ESpritz,[35] 2 versions of DisEMBL,[36] JRONN,[37] PONDR® VSL2B,[28,38] and GlobPlot.[39] MobiDB also has manually curated annotations related to protein function and structure derived from UniProt[24] and DisProt,[40] as well as from Pfam[41] and PDB.[42]

Complementary disorder evaluations together with important disorder-related functional information were retrieved from the D^2P^2 database (http://d2p2.pro/),[43] which is a database of predicted disorder for a large library of proteins from completely sequenced genomes.[43] D^2P^2 database uses outputs of IUPred,[29] PONDR® VLXT,[21] PrDOS,[44] PONDR® VSL2B,[28,38] PV2,[43] and ESpritz.[35] The database is further supplemented by data concerning location of various curated posttranslational modifications and predicted disorder-based protein binding sites.

Additional functional information for these proteins was retrieved using Search Tool for the Retrieval of Interacting Genes; STRING, http://string-db.org/, which generates a network of predicted associations based on predicted and experimentally-validated information on the interaction partners of a protein of interest.[45] In the corresponding network, the nodes correspond to proteins, whereas the edges show predicted or known functional associations. Seven types of evidence are used to build the corresponding network, where they are indicated by the differently colored lines: a green line represents neighborhood evidence; a red line – the presence of fusion evidence; a purple line – experimental evidence; a blue line – co-occurrence evidence; a light blue line – database evidence; a yellow line – text mining evidence; and a black line – co-expression evidence.[45] In our analysis, the most stringent criteria were used for selection of interacting proteins by choosing the highest cut-off of 0.9 as the minimal required confidence level.

Interactability of 7 major subunits of human hemoglobin was further evaluated by the APID (Agile Protein Interactomes DataServer) platform (http://apid.dep.usal.es).[46] APID contains information on 90,379 distinct proteins from more than 400 organisms (including *Homo sapiens*) and on the 678,441 singular protein-protein interactions. For each protein–protein interaction (PPI) the server provides currently reported information about its experimental validation. For each protein, APID unifies PPIs found in 5 major primary databases of molecular interactions, such as BioGRID,[47] Database of Interacting Proteins (DIP),[48] Human Protein Reference Database (HPRD),[49] IntAct,[50] and the Molecular Interaction (MINT) database,[51] as well as from the BioPlex (biophysical interactions of ORFeome-based complexes)[52] and from the protein databank (PDB) entries of protein complexes.[53] This server provides a simple way to evaluate the interactability of individual proteins in a given dataset and also allows researchers to create a specific protein-protein interaction network in which proteins from the query data set are engaged.

Results and discussion

Intrinsic disorder in the individual subunits of human hemoglobin

Hemoglobin in *Homo sapiens* is made of many different subunits that change during the development of the human. When a human is an adult, the hemoglobin protein is made of 2 α- subunits and 2 β-subunits. The mutation leading to the sickle cell anemia occurs in the N-terminal region of β-subunit. Before hemoglobin is able to develop α-subunits, it must have "combinations of ζ- with ε- or γ-subunits to form embryonic hemoglobins."[54] Their order of expression is determined by their relative positions on the gene, "i.e., $\zeta \rightarrow \alpha$ (2 copies) on chromosome 16 and $\varepsilon \rightarrow \gamma$ (2 copies) $\rightarrow \delta \rightarrow \beta$ on chromosome 11."[54] During normal development, the embryo is normally $\zeta_2\gamma_2$, $\zeta_2\varepsilon_2$, or $\alpha_2\varepsilon_2$, the fetus is typically $\alpha_2\gamma_2$, and, finally, the adult stage hemoglobin consists of either $\alpha_2\beta_2$ or $\alpha_2\delta_2$.[54] Since the hemoglobin is a hetero-tetramer consisting of several combinations of these 7 different types of subunits (α, β, $\gamma1$, $\gamma2$, δ, ε, and ζ), in order to understand the potential role of intrinsic disorder in hemoglobin function, the predisposition of all these subunits to intrinsic disorder were evaluated. To this end, protein sequences of each subunit were retrieved

from UniProtKB and utilized in the disorder analysis by several commonly used predictors, such as PONDR-FIT, PONDR® VLXT, PONDR® VSL2, and PONDR® VL3,[21,26-28,55,56] as well as the IUPred web server.[29] Although hemoglobin has more subunits (such as μ and θ), they were ignored in our study, because the majority of hemoglobin development relies on varying the combinations of the 7 main subunits: α, β, $\gamma1$, $\gamma2$, δ, ε, and ζ.

Clustal Omega was used to run a multiple sequence alignment of all 7 subunits, including both isoforms of the γ-subunit (see Fig. 1A). The multiple sequence alignment revealed that all sequences share 37 identical positions and 56 similar positions, considering that every sequence is between 142 and 147 amino acids long. The percent identity was 24.8%, which shows that the subunits of hemoglobin have a relatively low level of evolutionary conservation. The γ genes are duplicated: one codes for glycine (Glyγ) and the other for alanine (Alaγ) at residue 136, giving rise to 2 kinds of γ chains.[57] The phylogenetic tree showed that at first there was a divergence between the α- and ζ-subunits from the others (see Fig. 1B). This would make sense because they bind to all of the other hemoglobin subunits through the development of a human. Another divergence emerged leading to the separation of $\gamma1$-, $\gamma2$-, and ε-subunits from the β- and δ-subunits. This would probably involve the fact that the β- and δ-subunits bind to the α-subunits in adulthood, whereas the $\gamma1$-, $\gamma2$-, and ε-subunits bind to α- or ζ-subunits during embryonic development. Finally, there was a divergence between the ε- and $\gamma1/\gamma2$-subunits and eventually the divergence between the $\gamma1$- and $\gamma2$-subunits.

Figure 2 and Table 1 represent results of the disorder predisposition analysis in 7 major subunits of human hemoglobin and show that although these protomers are predicted to be mostly ordered, all of them are characterized by the presence of disordered regions (i.e., regions with the disorder scores above 0.5) or flexible regions (i.e., regions with the disorder scores above 0.2). Figure 2A-2G also indicate that despite low level of sequence similarity, different hemoglobin subunits are characterized by rather similar disorder profiles, which can be described as a specific pattern in disorder/flexibility distribution within the amino acid sequences observed for all 7 hemoglobin protomers: disordered N-tail (residues 1–7/9) – flexible/
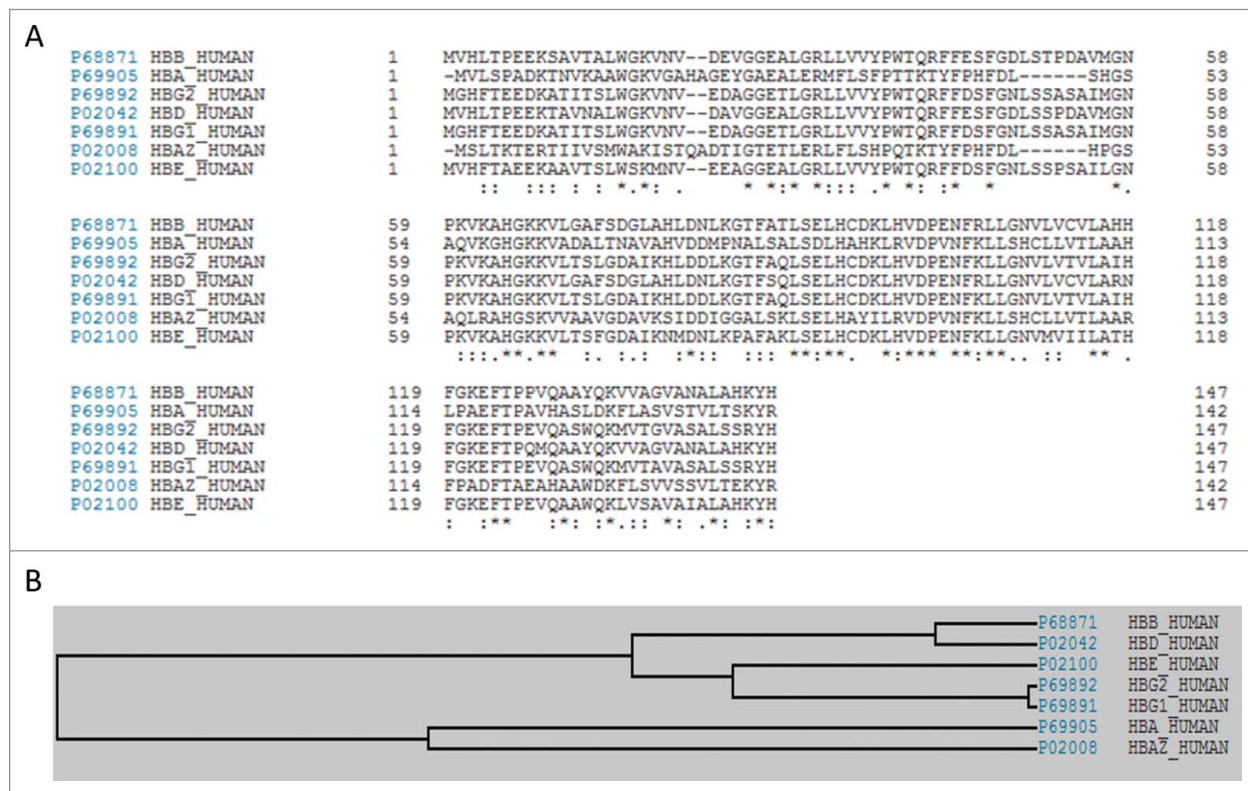

Figure 1. Evolutionary conservation of the major subunits of human hemoglobin. (A) Multiple sequence alignment of 7 major subunits of human hemoglobin. (B) Phylogenetic tree of 7 major subunits of human hemoglobin.
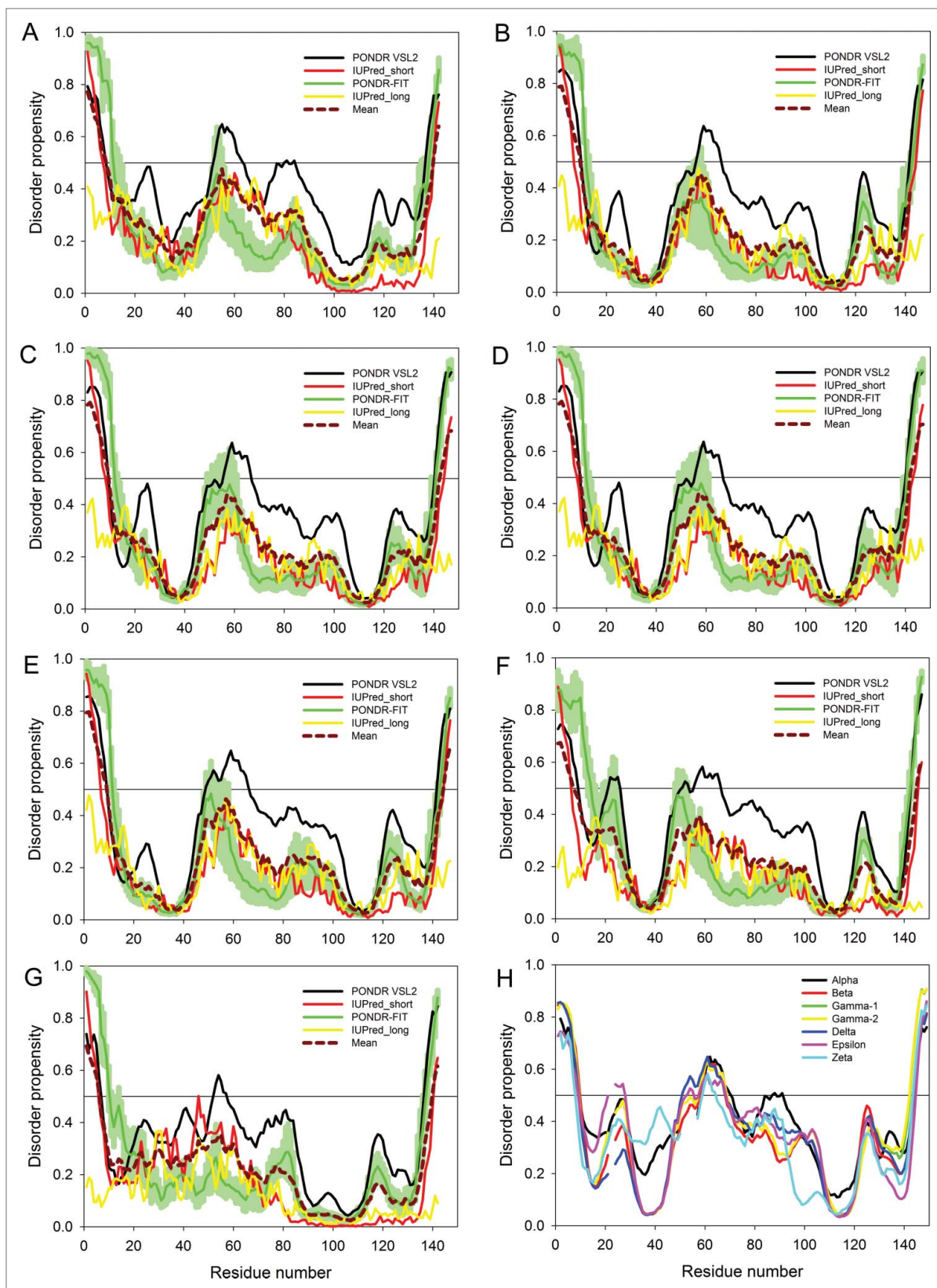
Figure 2. (For figure legend, see page 6.)

disordered region (residues 18/19–28/34) – disordered region (residues 49/56–60/66) combined with the long flexible region (residues 56/66–88/105) – flexible region (115/120–129/137) – disordered C-tail (residues 138/141–142/147). This observation is further illustrated by Fig. 2H, where the aligned PONDR® VSL2 profiles are shown. The fact that evolutionary diverged proteins are characterized by very similar disorder profiles suggests that the conserved peculiarities of the sequence distribution of predisposition for disorder/flexibility may have some functional implementations.

To further illustrate abundance and functionality of intrinsic disorder in various subunits of human hemoglobin (Hb), Fig. 3 represents the outputs of the D^2P^2 platform (http://d2p2.pro/),[43] which represents the disorder predisposition of a query protein by showing regions predicted to be disordered by IUPred,[29] PONDR® VLXT,[21] PrDOS,[44] PONDR® VSL2B,[28,38] PV2,[43] and ESpritz[35] and also shows location of various posttranslational modifications (PTMs). Since no D^2P^2 output is available for the $\gamma 1$-subunit (UniProt ID: P69891) as of yet, Fig. 3 represents these data for 6 human Hb subunits: α (UniProt ID: P69905), β (UniProt ID: P68871), $\gamma 2$ (UniProt ID: P69892), δ (UniProt ID: P02042), ε (UniProt ID: P02100), and ζ (UniProt ID: P02008). It is seen that in agreement with the data shown in Fig. 2, D^2P^2 profiles of all the human Hb protomers contain regions of intrinsic disorder (i.e., regions with disorder scores exceeding 0.5 threshold). Furthermore, all these proteins are heavily decorated with various PTMs, such as phosphorylation, acetylation, glycosylation, ubiquitination, and nitrosylation. Curiously, although all 6 subunits have multiple phosphorylation and acetylation sites, different Hb subunits show clear preference for some specific PTMs. For example, glycosylation and nitrosylation sites are present only in α- and β-subunits, whereas other subunits do not have these PTMs; ubiquitination sites are absent in β- and δ-subunits, whereas other subunits have 3 to 10 ubiquitination

sites (see Fig. 3). It is also clear that the majority of all these numerous PTMs are preferentially located within disordered or flexible regions. All this clearly indicates that conserved regions of intrinsic disorder or conformational flexibility have several important functions for these proteins.

Correlating intrinsic disorder propensities with known structural information

For the tetrameric hemoglobin (Hb) to function, the existence of at least 2 interconvertible quaternary forms with different affinities to oxygen/carbon monoxide is needed, the deoxy-HB or T-state, which is the tense, ligand-free form and the oxy-Hb or R-state, which is the relaxed, ligand bound form in complex with oxygen or carbon monoxide.[58,59] The presence of these 2 forms was supported by first crystallographic studies of human adult Hb, based on which the T-R transition was attributed to significant changes in the Hb quaternary structure, with the tertiary structures of each subunit being nearly identical in the 2 forms and with almost unchanged $\alpha_1\beta_1$ dimer.[60] It was pointed out that the substantial difference in the quaternary structure associated with the T-R transition is caused mainly by changes in the number and identity of groups involved in the pairwise interactions across the $\alpha_1\beta_2 – \alpha_2\beta_1$ interfaces, whereas there is no noticeable changes in the pairwise interactions across the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ interfaces.[61] More detailed structural analyses of ligand-bound Hb structures generated by altering its crystallization conditions revealed that several other carbonmonoxy Hb (COHbA) forms can be present, such as R2,[62] RR2,[63] and R3.[63] It was pointed out that all these forms are characterized by significantly different quaternary structures, especially at their $\alpha_1\beta_2$ dimer interfaces.[63] Curiously, although it was pointed out that tertiary structures of the α- and β-protomers in human adult Hb are not significantly altered by the T → R, T → R → R2, T → RR2, and T → R3 transitions, some of the regions of β-subunit in

Figure 2. (see previous page) Multifactorial disorder profiles of 7 major subunits of human Hb. Evaluating intrinsic disorder propensity by series of per-residue disorder predictors. Disorder profiles generated by PONDR® VSL2, IUPred_short, PONDR® FIT, and IUPred_long are shown by black, red, green, and yellow lines, respectively. Dark red dashed line shows the mean disorder propensity calculated by averaging disorder profiles of individual predictors. Light green shadow around the PONDR® FIT curves shows error distribution. In these analyses, the predicted intrinsic disorder scores above 0.5 are considered to correspond to the disordered residues/regions, whereas regions with the disorder scores between 0.2 and 0.5 are considered flexible. Disorder profiles for the α-, β-, $\gamma 1$-, $\gamma 2$-, δ-, ε-, and ζ-subunits are shown in plots A, B, C, D, E, F, and G, respectively. H. Aligned PONDR® VSL2-based disorder profiles for the α-, β-, $\gamma 1$-, $\gamma 2$-, δ-, ε-, and ζ-subunits of human Hb. Outputs for the individual subunits were aligned using results of multiple sequence alignment shown in Fig. 1A. Gaps in the individual disorder profiles correspond to the deletions.

Table 1. Characterization of the major subunits of human hemoglobin.

Protein name and UniProt ID	Length (Molecular mass, Da)	pI	Content of the predicted disordered residues (CPDR, %)						N_{APID}
			VSL2	FIT	IUPred_s	IUPred_l[a]	Mean	MobiDB	
α-Subunit P69905	142 (15257.55)	8.72	20.42	11.26	7.04	0.209 ± 0.010	7.75	8.45	73
β-Subunit P68871	147 (15998.41)	6.74	17.69	10.88	6.12	0.179 ± 0.008	8.16	6.80	54
β-Subunit E6V P68871	147 (15968.43)	7.13	17.01	10.88	5.44	0.162 ± 0.008	6.12	N/A	N/A
$\gamma1$-Subunit P69891	147 (16140.47)	6.64	18.37	12.24	7.48	0.183 ± 0.008	9.52	9.52	12
$\gamma2$-Subunit P69892	147 (16126.44)	6.64	18.37	12.24	8.16	0.190 ± 0.008	10.20	9.52	14
δ-Subunit P02042	147 (16055.48)	7.84	22.45	10.88	6.12	0.182 ± 0.009	8.16	5.44	23
ε-Subunit P02100	147 (16202.82)	8.67	21.77	12.93	5.44	0.146 ± 0.008	6.80	6.12	8
ζ-Subunit P02008	142 (15637.02)	7.94	12.68	10.56	5.63	0.124 ± 0.007	6.34	5.63	29

the R3 form, which is and $\alpha_1\beta_1$ dimer, as well as in RR2 form, which is $\alpha_1\beta_1\alpha_2\beta_2$ Hb tetramer were characterized by the increased mobility as evidenced by their high B-factors.[63] The mobile regions in the RR2 and R3 forms are the N-terminus (residues 1—3), E helix—EF corner—F helix—FG corner region (residues 58—103), and C-terminus (residues 139—146); i.e., the regions which were predicted to be disordered and/or flexible in our analysis. In the RR2 and R3 form, these mobile regions are exposed to solvent and are involved in the formation of the β-cleft or serve as the residues around the β-heme.[63] Solution NMR analysis of human COHbA revealed that this protein can exist in multiple conformations, whose structures were similar to the R quaternary structure.[60] Fig. 4 represents a structural snap-shot of human COHbA in solution that includes 20 conformers for each COHbA subunit and shows that this tetrameric protein is characterized by noticeable structural variations in solution.[60]

Although amino acid sequence is considered the most reliable source for predicting intrinsically disordered regions within a protein, some useful information related to protein flexibility can also be obtained by predicting secondary and tertiary structure of a target protein. The SWISS-MODEL Repository was used to create models of 7 major subunits of human Hb in order to look for the correlations between the predicted regions of intrinsic disorder and regions of structural flexibility in predicted 3D models. Although the structure prediction might not be the most reliable method, but it can still provide a useful 3-dimensional image of the protein and represent the distribution of threading energy throughout the entire molecule. When viewing protein structures generated by the SWISS-MODEL, dark blue regions indicate that the threading energy was low and that the corresponding residues were properly set in their positions (did not move), whereas red regions indicate that the threading

energy was high and that the corresponding protein region is considered entropic or unsettled in its environment. Furthermore, analysis of the 3-D structure of the protein could possibly predict intrinsically disordered regions because the structure binding-folding thermodynamics and kinetics, which are important for the efficiency of realizing biomolecular function, can be deduced from its global energy landscape topology.[64,65] Therefore, the IDPRs of the hemoglobin subunits could be further analyzed by the levels of threading energy detected by the SWISS-MODEL Repository protein structures, since intrinsic disorder is characterized by high entropy and lack of the defined structure.

Although SWISS-MODEL might not be an accurate predictor of intrinsic disorder, it still provides an accurate measure of the distribution of threading energy, which is essential to the biological function and defined structure of the query protein. According to the disorder predictions (see Figs. 2 and 3), the α-subunit has 4 IDPRs surrounded by *flexible* regions: residues 1–8(*9–33*), (*36–50*)51–63(*64–79*)80–81(*82–99*), and (*114–136*)137–142. Figure 5A and 5a shows that the most prominent red regions can be found around residues 41–48, 56–67, 82–102, and 136–142 of the α-subunit. Therefore, the red regions are either coincide, or overlap, or are located in close proximity to the predicted IDPRs, suggesting that the local predisposition for intrinsic disorder encoded in the amino acid sequence might cause lack of defined structure or high structural mobility of the corresponding regions.

Very similar situation is observed for the β-subunit, which is predicted to have 3 IDPRs and 4 flexible regions, residues 1–9(*10–13*), (*44–55*)56–66(*67–104*), and (*120–141*)142–147. Figure 5B and 5b shows that the most prominent red regions are around residues 37–46, 63–73, 87–100, and 142–147. Although it seems that the number of red regions in modeled structure is larger than the
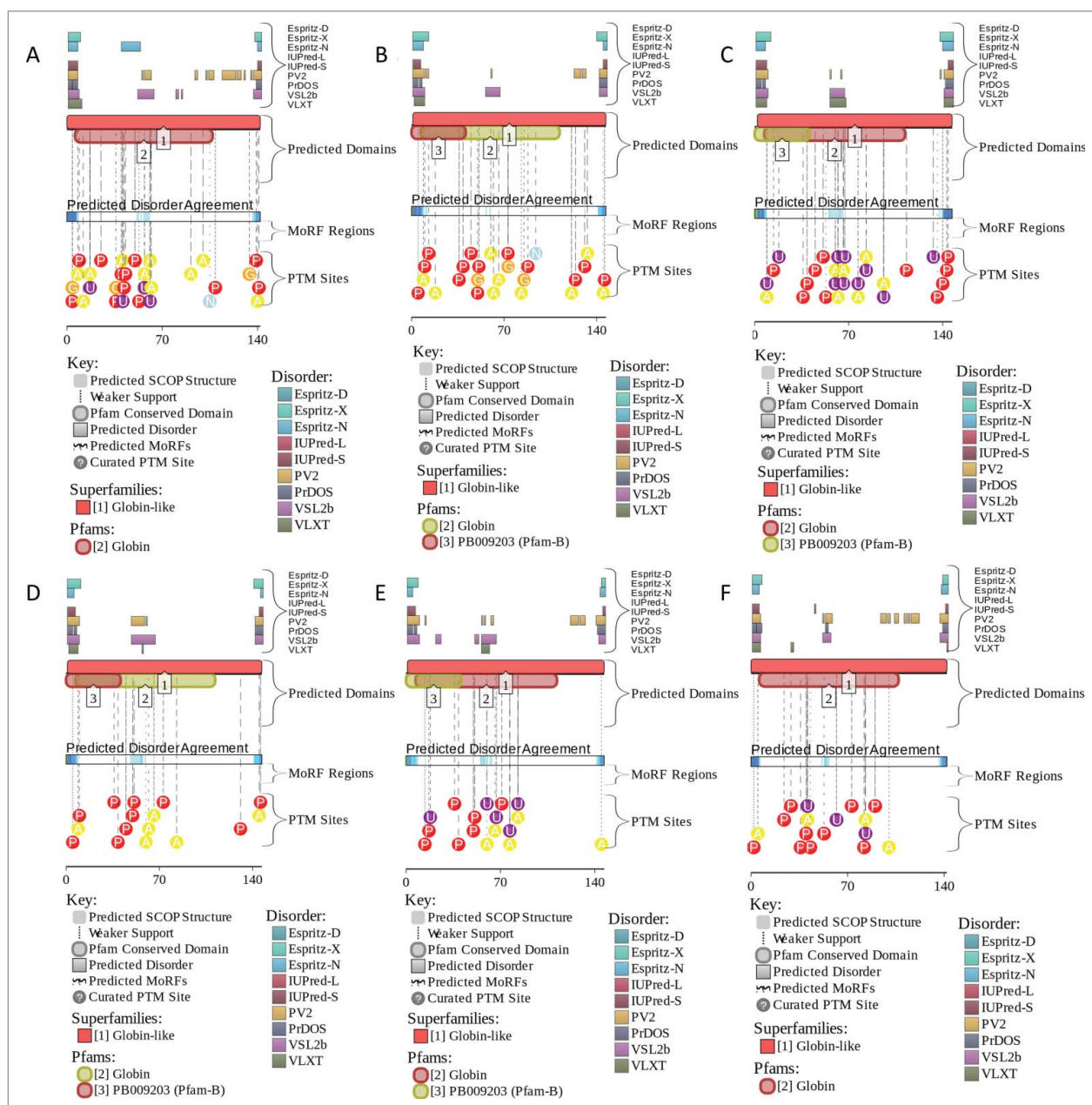
Figure 3. Intrinsic disorder propensity and some important disorder-related functional information generated by the D²P² database (http://d2p2.pro/)[43] for the major subunits of human Hb: (A) α-subunit; (B) β-subunit; (C) γ1-subunit; (D) γ2-subunit; (E) δ-subunit; (F) ε-subunit; and (G) ζ-subunit. Here, the green-and-white bar in the middle of the plot shows the predicted disorder agreement between 9 predictors, with green parts corresponding to disordered regions by consensus. Colored circles at the bottom of the each plot show location of various PTMs.

number of predicted IDPRs, one should keep in mind that some of the predicted IDPRs have rather long flexible tails (i.e., regions with the disorder scores above 0.2). For example, regions 44–55 and 67–104 preceding and following the second IDPR are predicted to be flexible. On the other hand, this second IDPR (residues 56–66) is surrounded by red regions (residues 37–46 and 63–73) suggesting that local intrinsic disorder might cause lack of defined structure to surrounding regions.

Both isoforms of the γ-subunit have 3 IDPRs and 4 *flexible* regions each, residues 1–9(*10–13*), (*44–55*) 56–66(*67–107*), and (*121–140*)141–147. The SWISS-MODEL structure for the γ1-subunit has a few prominent red regions, the most prominent of which are around the residues 38–47, 64–72, 88–107 and 142–147 (see Fig. 5C and 5c). The SWISS-MODEL structure for the γ2-subunit has many violet and weak red regions, with the most prominent red
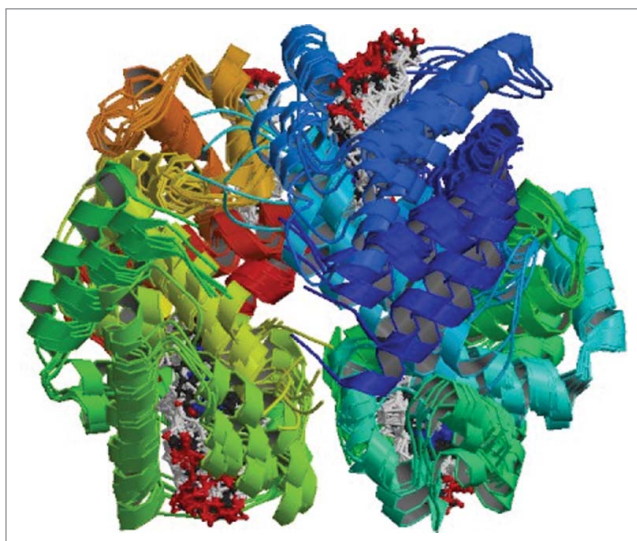
Figure 4. NMR solution structure of the human adult Hb (PDB ID: 2M6Z; ref. 60). A superposition of 20 conformers is shown to illustrate the presence of noticeable conformational dynamics in this protein.

regions being roughly positioned at residues 38–43, 93–98, and 145–147 (see Fig. 5D and 5d). Both isoforms of the γ-subunit do not have many prominent red regions. Their regions with highest energy levels are in similar locations, although the γ1-subunit has much more pronounced red coloring between residues 60 and 75. Interestingly, similar to the α- and β-subunits, the N-terminal IDPRs of the γ1- and γ2-subunits were not the disordered or flexible regions in the corresponding modeled structures. On the other hand, the C-terminal IDPRs was found highly flexible in all these structures. Compared to the α- and β-subunits, the γ1-subunit seems quite similar in distribution of its most prominently red regions. The γ2-subunit might not share the exact residue positions for prominent red areas, but both isoforms of the γ-subunits showed roughly similar distribution of red regions and the γ2-subunit has much less flexibility than the α-, β-, and γ1-subunits.

The δ-subunit is predicted to have IDPRs and *flexible* regions at positions 1–9(*10–23*), (*44–48*)49–66(*67–104*), and (*120–141*)142–147. Figures 5E and 5e shows that the most prominent red regions of this subunit are around the residues 37–46, 88–100, and 143–147. Overall, the δ-subunit has roughly the same red regions as the α- and β-subunits, except that the red regions are less prominent and that the region around 58–72 is either violet or blue. This time there are 2 prominently red region that are adjacent to the IDPRs.

The ε-subunit has several prominent IDPRs surrounded by *flexible* regions located at residues 1–9 (*10–21*)22–25(*26–29*), (*46–50*)51–66(*67–105*), and (*141–142*)143–147. The SWISS-MODEL structure for the ε-subunit has most prominent red regions around the residues 38–46, 64–72, 90–107, and 142–146. Once again, the idea that high intrinsic disorder propensity affects the threading energy of its surrounding regions is seen. The position of the most prominent red regions in the ε-subunit resembles those in α-, β-, and γ1-subunits.

Finally, the ζ-subunit has 3 disordered regions and 4 *flexible* regions around the residues 1–6(*7–11*), (*16–51*)52–57(*58–87*), and (*134–137*)138–142. The strongest red regions in this subunit are roughly 40–47, 59–66, 84–102, and 133–142. Besides N-terminal disordered tail, other IDPRs are rather accurately reflected in the predicted 3D structure. The distribution of flexible regions in the ζ-subunit is rather similar to those of α-, β-, ε-, and γ1-subunits.

Posttranslational modifications of the hemoglobin subunits

Posttranslational modifications (PTMs) are covalent modifications of proteins that are typically catalyzed by specific enzymes after the translation of a polypeptide chain. PTMs are many and have various functional implementations, such as functional regulation of proteins, control of cellular localization, or targeting proteins for proteolytic cleavage. The most common PTMs are phosphorylation, glycosylation, acetylation, nitrosylation, ubiquitination, and many others. It is also known that the disordered domains/regions of many may have several posttranslational modification sites, and phosphorylation[66] and many other enzymatically catalyzed PTMs are preferentially located within the IDPRs.[67]

Figure 3 shows that all hemoglobin subunits analyzed in this study (except for γ1-subunit for which D^2P^2 is not available as of yet) have a large number of various PTMs (see Fig. 3). In addition to the PTMs visualized by D^2P^2, information on the experimentally validated PTMs of 7 major Hb subunits was extracted from the UniProtKB. This analysis revealed that the α-subunit of human Hb is phosphorylated at numerous sites, including serine residues at positions 4, 36, 50, 103, 125, 132, and 139; threonine residues at positions 9, 109, 135, and 138; and a tyrosine residue at
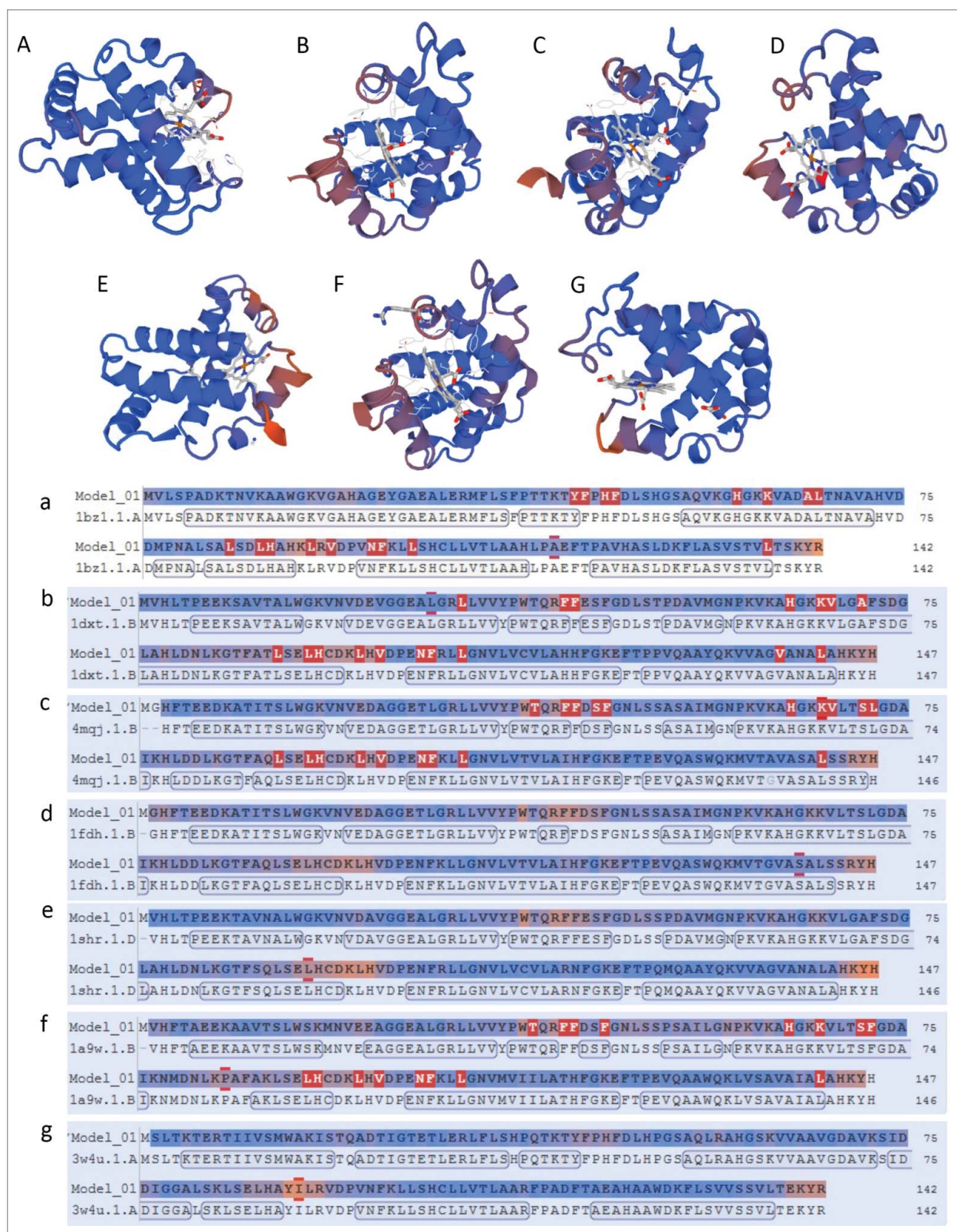
Figure 5. SWISS-MODEL protein structures (A, B, C, D, E, F, and G) and corresponding threading energies (a, b, c, d, e, f, and g) of 7 major subunits of human Hb: A, a. α-Subunit; B, b. β-Subunit; C, c. γ1-Subunit; D, d. γ2-Subunit; E, e. δ-Subunit; F, f. ε-Subunit; and G, g. ζ-Subunit.

position 25. The second most frequent posttranslational modification in this subunit is glycosylation, which is found at positions 8, 17, 41, and 62. The lysine residues at positions 8, 12, 17, and 41 are N6-succinylated. Finally, the lysine residue at position 17 is N6-acetylated. Since the disorder-*flexibility* distribution within the α-subunit represents the following pattern: 1–8(*9–33*) – (*36–50*)51–63(*64–79*)80–81(*82–99*) – (*114–136*)137–142, all the PTMs are located within disordered or flexible regions.

The β-subunit of human hemoglobin has posttranslational modifications at positions 2, 9, 10, 13, 18, 45, 51, 60, 67, 83, 88, 94, 121, and 145. The amino acid valine at position 2 can be N-acetylated, glycosylated, and pyruvic acid iminylated. The β-subunit is glycosylated at positions 9, 18, 67, 121, and 145 and has several phosphorylated sites, such as the serine residues at positions 10 and 45 and the threonine residues at positions 13, 51, and 88. The lysine residues at positions 60, 83, and 145 are N6-acetylated. Finally, the cysteine residue at position 94 is S-nitrosylated. All PTMs are located within the disordered or flexible regions in this subunit, which is characterized by the following disorder/*flexibility* profile: 1–9(*10–13*) – (*44–55*)56–66(*67–104*) – (*120–141*)142–147.

According to UniProtKB, only 2 PTM types are found in the hemoglobin γ1- and γ2-subunits, N-acetylation and phosphorylation. In both isoforms, the N-acetylation takes place at the Gly2 position, whereas phosphorylation occurs at serine residues 45, 51, 53, 140, 143, and 144. The disorder/*flexibility* profiles of the hemoglobin γ1- and γ2-subunits are identical: 1–9(*10–13*) – (*44–55*)56–66(*67–107*) – (*121–140*)141–147, indicating that all PTMs happen either in disordered or flexible regions.

The δ-subunit has 2 phosphorylation sites (residues Ser45 and Ser51), 3 N6-acetylation sites (residues Ly60, Lys83, and Lys145) and one S-nitrosylation site (Cys94). An additional modification, N-acetylation of the Ala2 is found in the Niigata variant of this subunit that is present in Japanese population.[68] Since the disorder-*flexibility* pattern of this subunit can be described as 1–9(*10–23*) – (*44–48*)49–66(*67–104*) – (*120–141*)142–147, all these PTMs are located within disordered or flexible regions.

The hemoglobin ε-subunit has 3 phosphorylated amino acid residues: 2 serine residues at positions 45 and 51 and threonine at position 124. It also has 2

N6-succinylated lysine residues at positions 18 and 60, and one S-nitrosylated cysteine residue at position 94. Several IDPRs surrounded by flexible regions represent the disorder/*flexibility* profile of this subunit 1–9 (*10–21*)22–25(*26–29*) – (*46–50*)51–66(*67–105*) – (*141–142*)143–147, indicating that all PTMs sites of this subunit are located within disordered or flexible regions.

The hemoglobin ζ-subunit is characterized by the presence of one N-acetylation site at the position Gly2 and 4 phosphorylation sites at positions Thr29, Ser53, Ser73, and Ser82. The PONDR® VSL2-based disorder/*flexibility* profile of ζ-subunit (1–6(*7–11*) – (*16–51*)52–57(*58–87*) – (*134–137*)138–142) indicates that similar to all other Hb subunits analyzed in this study, all these PTMs sites tend to be located in disordered or flexible regions of ζ-subunit.

Interactions with protein partners

Besides being involved in interaction with each other, the major subunits of human hemoglobin have multiple interactions with a wide variety of other proteins. The corresponding interactomes were discovered through the STRING platform, which generates protein-protein interaction (PPI) networks to determine the functional repertoire of a query protein. Construction and analysis of these interactomes are important, since such PPIs could provide important clues on the biological activities of each hemoglobin subunit.

According to the STRING-based analysis, the human Hb α-subunit (HBA1) has only 5 interactions when the highest confidence of 0.9 is used. Figure 6A shows that 3 of the HBA1 interactors are other hemoglobin subunits, hemoglobin β (HBB), hemoglobin epsilon 1 (HBE1), and hemoglobin α 2 (HBA2). This Hb subunit can also interact with the α hemoglobin stabilizing protein (AHSP) and haptoglobin (HP). Curiously, both "out of family" interactions are with proteins that are involved in regulation of hemoglobin proteostasis. In fact, AHSP is a chaperone that binds to free Hb α-subunit and protects this subunit from the harmful aggregation during normal erythroid cell development. HP, which captures and interacts with free plasma hemoglobin, leading to the effective clearance of the hemoglobin/haptoglobin complexes by the macrophages. This is an important part of the hepatic recycling of heme iron that prevents potential kidney damage associated with the hemolysis leading to the
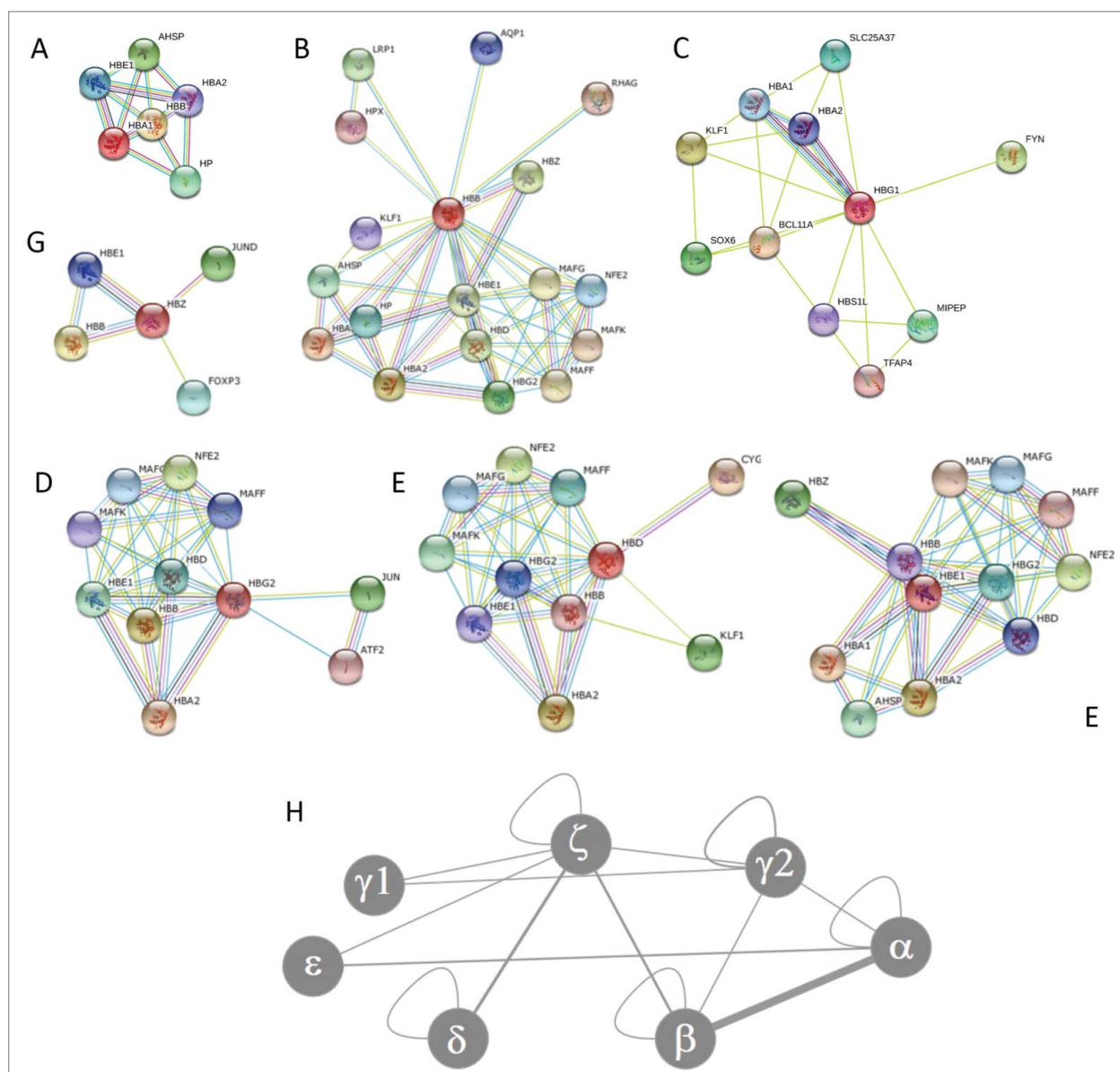
Figure 6. Analysis of the interactivity of major subunits of human Hb: (A) α-subunit; (B) β-subunit; (C) γ1-subunit; (D) γ2-subunit; (E) δ-subunit; (F) ε-subunit; and (G) ζ-subunit by STRING computational platform that produces the network of predicted associations for a particular group of proteins.[45] (H) The output of the APID platform for the 7 major subunits of human hemoglobin.

accumulation of hemoglobin in the kidney and secretion to the urine.

As shown in Fig. 6B, the human Hb β-subunit (HBB) interacts with 17 different proteins, 6 of which are hemoglobin subunits HBA1 (hemoglobin α 1), HBA2 (hemoglobin α 2), HBZ (hemoglobin zeta), HBD (hemoglobin delta), HBE1 (hemoglobin epsilon 1), and HBG2 (hemoglobin gamma G). Similar to HBA1, the β-subunit also interacts with the AHSP chaperone and HP. Other proteins that interact with the β-subunit include the 3 different homologs of v-maf musculoaponeurotic fibrosarcoma oncogene, F

(MAFF), K (MAFK), and J (MAFJ), all of which are transcription factors that act as transcriptional activators or repressors involved in embryonic lens fiber cell development. There are also other transcriptional factors that interact with HBB, such as the Kruppel-like factor (KLF1), which is a DNA-binding protein that is involved in gene expression regulation, and nuclear factory erythroid 2 (NFE2), which is involved in megakaryocyte production. Hemopexin is another important binding partner of the Hb β-subunit needed to avoid oxidative damage, since this protein is responsible for inhibiting the oxidative activity of low-

affinity hemoglobin that is released by erythrocytes. The Rh-associated glycoprotein (RHAG), another member of the HBB interactome, is an ammonia transporter protein. Finally, HBB interacts with aquaporin 1 (AQP1), which is a water channel found in the plasma membranes of certain regions of nephrons, and low density lipoprotein receptor-related protein 1 (LRP1), which is a receptor that is responsible for the process of receptor-mediated endocytosis. All of these interactions define the diverse functionality of the Hb β-subunit.

Currently available information on the Hb γ1-subunit (HBG1) interactivity is rather limited. In fact, STRING did not generate a PPI network for this protein when the highest confidence of 0.9 was used. Relaxing the confidence level to "high confidence" of 0.7 allowed obtaining the PPI network containing 4 interaction partners (SRY (sex determining region Y)-box 6 (SOX6), Kruppel-like factor 1 (KLF1), B-cell CLL/lymphoma 11A (BCL11A), and FYN oncogene), none of which were other Hb subunits. Since the Hb γ1-subunit is known to be engaged in the formation of a heterotetramer with the Hb α-subunit needed for the formation of the fetal/embryonic Hb, we further relaxed the confidence level to "medium confidence" of 0.4. The resulting PPI network is shown in Fig. 6C, where in addition to the aforementioned SOX6, KLF1, BCL11A, and FYN and the hemoglobin α-subunits HBA1 and HBA2, the Hb γ1-subunit is shown to interact with solute carrier family 25 (mitochondrial iron transporter), member 37 (SLC25A37), mitochondrial intermediate peptidase (MIPEP), transcription factor AP-4 (TFAP4 or activating enhancer binding protein 4), and HBS1-like protein (HBS1L).

The Hb γ2-subunit (HBG2) interacts with other Hb subunits, HBA2, HBB, HBE1, and HBD. Similar to HBB, HBG2 also interacts with MAFG, MAFF, MAFK, and NFE2 (see Fig. 6D). It also interacts with 2 transcription factors: jun proto-oncogene (JUN) and activating transcription factor 2 (ATF2).

Figure 6E shows that the hemoglobin δ-subunit interacts with 4 other hemoglobin subunits, HBA2, HBG2, HBE1, and HBB. It also can interact with MAFK, MAFF, MAFG, NFE2, and KLF1. In addition, the δ-subunit interacts with cytoglobin, which is a globin molecule that participates in the prevention of the oxidative stress and scavenges reactive oxygen species and nitric oxide. The cytoglobin, if altered, could lead to enhanced oxidative stress

causing enhanced cellular damage of the erythrocytes.

The Hb ε-subunit (HBE1), which is the other fetal globin, interacts with hemoglobin subunits HBA1, HBA2, HBZ, HBG2, HBD, and HBB. Similar to HBB, it can also interact with NFE2, AHSP, MAFG, MAFF, and MAFK (see Fig. 6F); i.e., proteins needed for hemoglobin stability, oxygen affinity, transport, and endocytosis, among other important functions.

The fetal hemoglobin molecule, the human Hb ζ-subunit, interacts with only 2 of the other hemoglobin subunits, HBB and HBE1. It also interacts with JUND and forkhead box P3 (FOXP3), which is a protein that regulates the development of regulatory T cells (see Fig. 6G). Although the ζ-subunit seems to have a limited interactome, it still plays an important role in embryonic and fetal development.

Table 1 represents the results of the application of the APID server for evaluation of the interactivity of the 7 major subunits of human hemoglobin and clearly show that each of these proteins is engaged in multiple protein-protein interactions (PPIs) (see Table 1). In fact, the number of PPIs ranges from 8 to 73, with the majority of these proteins being able to interact with more than 10 partners each. This observation suggests that the hemoglobin subunits can considered as hub proteins. Obviously, hemoglobin subunits are involved in interactions with each other. To illustrate this inter-hemoglobin interactivity, we used the ability of the APID web server (http://apid.dep.usal.es) to build a specific PPI network between proteins included in a query list.[46] Fig. 6H represents the results of application of this tool to 7 major subunits of human hemoglobin and shows that according to their inter-hemoglobin interactivity, these subunit can be arranged in the following order: ζ (5) > γ2 (4) > α (3) = β (3) > γ1 (2) = ε (2) > δ (1). This APID-based analysis clearly shows that both internal (interactions between the subunits of hemoglobin) and external connectivities (interaction with other proteins) are high for the major hemoglobin subunits.

Sickle cell anemia-related mutation and intrinsic disorder

The sickle cell anemia is caused by single point mutation (an adenine-thymine substitution) that affects the 6th codon of the *HBB* gene encoding the hemoglobin subunit β (β-globin). As a result, a disorder-

promoting charged glutamic acid is substituted by the order-promoting hydrophobic valine residue (β-6 glutamic acid is changed to β-6 valine). One should keep in mind that sequence numbering discussed here is related to the mature form of the protein with the removed initiator methionine, whereas all computational studies discussed in this work were conducted with the hB subunits containing their initiator methionines. Figure 7 represents the expected outcomes of this Glu6/7-Val substitution that modulate the intrinsic disorder propensity of the β-subunit in the form of the "disorder difference spectra." The corresponding plot was calculated as a simple difference between the disorder curves calculated by different disorder predictors for the mutant and wild type β-subunit. Obviously, in this presentation, negative peaks correspond to the regions with the locally decreased intrinsic disorder propensity due to the mutation, whereas positive peaks show regions where the local intrinsic disorder propensity is increased due to the mutation. Figure 7A shows that Glu6/7-Val substitution has a profound effect on the intrinsic disorder propensity of the N-terminal tail of the β-subunit (the sickle cell–related mutant is characterized by the noticeable decrease in local intrinsic disorder propensity), providing support to the idea that specific distribution of order and disorder is needed for the appropriate functioning of this protein.

Let us now consider how this point mutation affects some of the functional and structural features discussed in this article. As it was mentioned, normal Hb is a heterotetramer that contains 2 pairs of related polypeptide chains, where one chain of each pair is α-type chain (which is either ζ-subunit in the early embryogenesis or α-subunit present in all human Hbs encountered after early embryogenesis) and the other is non-α-type subunit (β, γ, or δ). The non-α chains varies depending on the stage of development, with the fetal Hb being of $\alpha_2\gamma_2$ type due to the presence of the γ-subunits, and the normal adult Hb being $\alpha_2\beta_2$ tetramers that include the β-chain. Additionally, there is the minor adult Hb form $\alpha_2\delta_2$ that accounts for $\sim$2.5% of the Hb of normal adults and contains the δ-subunit as non-α-chain. Therefore, during normal development, composition of hemoglobin undergoes a set of changes, with the embryonic Hb being normally $\zeta_2\gamma_2$, $\zeta_2\varepsilon_2$, or $\alpha_2\varepsilon_2$, the fetus typically having an Hb of $\alpha_2\gamma_2$ type, and, finally, the adult stage hemoglobin consisting of either $\alpha_2\beta_2$ or $\alpha_2\delta_2$.[54] As it follows from

this description, sickle cell anemia, that is caused by the mutation in the β-subunit found in adult Hb only, represents a threat to the mutation carrier mostly after the birth (of course, pregnancy complications of a mother with this mutation and her baby cannot be ignored too).

The pathophysiological basis of the sickle cell anemia is the sickling process by which the HbS (hemoglobin S or sickle hemoglobin that normally carries oxygen and represents a complex containing a heme, 2 normal α-globin chains, and 2 abnormal β-globin chains) is converted to the semi-solid aggregated form once oxygen is unloaded to the tissues during the deoxygenation process. Here, the aggregated HbS distorts red blood cells, decreases their flexibility and eventually causes permanent red-cell damage as a result of the repeated deoxygenation cycles.[69] The sickling process is markedly accelerated when intracellular concentration of HbS is increased.[69] At the structural level, soluble tetramer-insoluble aggregate transition happens when in the deoxygenated form of Hb S, the β-6 valine becomes buried in a hydrophobic pocket on an adjacent β-globin chain, thereby joining the molecules together to eventually form insoluble polymers.[70] It is very likely that such targeted Val6 insertion to the hydrophobic pocket of the neighboring β-subunit became plausible due to the fact that this hydrophobic residue is located within the disordered N-terminal region, indicating the potential role of conformational flexibility and distorted charge-hydrophobicity pattern within the N-terminal region of the β-subunit in the pathogenesis of sickle cell disease.

Since the β-subunit was shown to be heavily decorated with various PTMs, it is logical to analyze the effect of the Glu6/7-Val substitution on the predisposition of this protein to be modified. However, because none of the known PTMs was directly associated with Glu6, we utilized a recently developed computational tool ModPred Windows OS User-interface version.[67] Here, the amino acid sequences of the wild type and mutant Hb β-subunit I were used to predict their predispositions for various PTMs (see Fig. 7B and 7C for the PTM profiles of the wild type and mutant β-subunit). This analysis revealed the presence of some mutation-induced changes in the predisposition of the residues of this protein for PTMs: the potential amidation site at Glu7, potential phosphorylation/amidation/
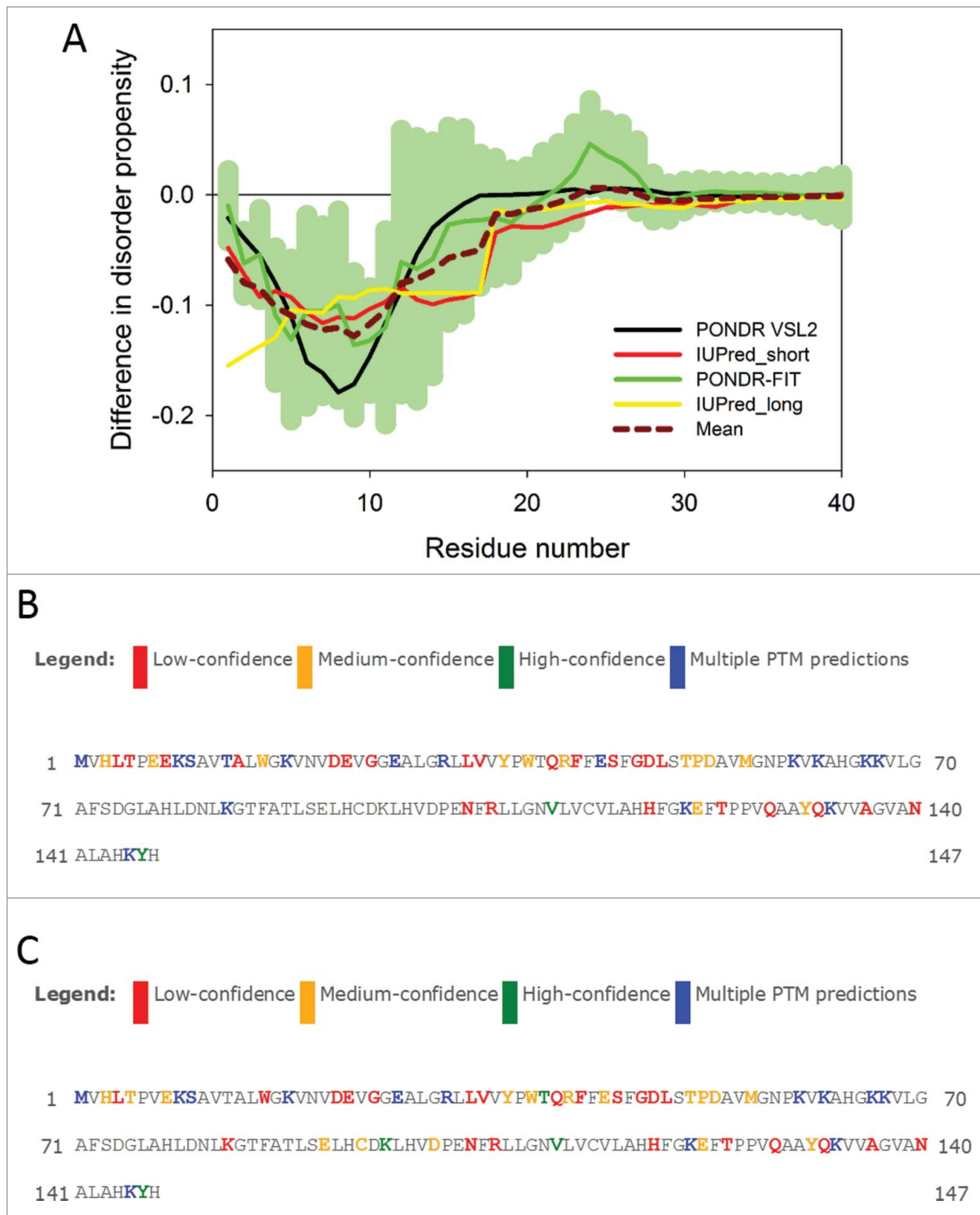
Figure 7. Effect of the sickle cell disease-associated mutation on local intrinsic disorder propensity and predisposition for PTMs of the β-chain of human Hb. (A) The "difference disorder spectrum" calculated as a simple difference between the per-residue disorder propensities evaluated for the mutant and wild type proteins by PONDR® VSL2, IUPred_short, PONDR® FIT, and IUPred_long are shown by black, red, green, and yellow lines, respectively. Dark red dashed line shows the mean disorder propensity calculated by averaging disorder profiles of individual predictors. Light green shadow around the PONDR® FIT curves shows distribution of errors. (B) The output of the ModPred predictor of potential PTMs[67] for the wild type β-chain of human Hb. (C) The output of the ModPred predictor of potential PTMs[67] for the sickle cell anemia-related mutant of the human Hb β-chain.

proteolytic cleavage site at Thr13, and potential proteolytic cleavage site at Ala14 were eliminated, whereas 3 new potential amidation sites were added at positions Glu91, Cys94, and Lys96. The probability of Lys83 to be amidated or methylated was downgraded to serve only as an amidation site. Glu8 was predicted to change its predisposition for PTM from low-confidence ADP-ribosylation in the wild type protein to high-confidence amidation in the sickle cell mutant, whereas Trp8 was predicted to change its predisposition for medium-confidence proteolytic cleavage to low-confidence amidation due to the mutation. These observations suggest that although sickle cell-causing mutation in the Hb β-subunit does affect the predisposition of this protein for various PTMs and not only at the mutation site *per se*, suggesting the potential effect of this mutation on the PTM-based regulation of the β-subunit functionality. In other words, since the Hb β-chain is the subunit directly involved in sickle cell anemia, the intrinsic disorder within N-terminal region of this protein might play a role in sickle cell disease.

Finally, the potential effect of sickle cell mutation on the Hb β-subunit interactivity should be pointed out. In fact, if mutant chain is characterized by altered capability to interact with its numerous binding partners, then various important biological processes (and not only oxygen binding) might be altered too. This is in line with the known fact that sickle cell RBCs are not able to transport or bind to many molecules due to their altered shape and loss of function.

Conclusions

Our analysis revealed that intrinsic disorder is present in all subunits of human hemoglobin, which are characterized by the remarkably similar disorder profiles despite their relatively low sequence similarities. Interestingly, the fetal subunits showed more/longer IDPRs and flexible regions than the adult hemoglobin subunits, suggesting higher need for intrinsic disorder in hemoglobin during the embryonic development. Sickle cell anemia-related mutation decreases the local intrinsic disorder propensity, changes the local predisposition for PTMs and likely affects the capability of the β-subunit to interact with its numerous protein partners.

Generation of 3D structural models for 7 major subunits of human Hb by SWISS-MODEL was a

reinforcement of the proposed predisposition of these chains for intrinsic disorder. Overall, there was a reasonable good correlation between threading energy levels in certain protein regions and their propensity for intrinsic disorder/flexibility. Although some of the IDPRs were not characterized by high entropy values in corresponding 3D structures, many of them were located between or adjacent to areas of high threading energy levels. These observations suggest that although some IDPRs themselves might not have high threading energy levels or unsettled residues, they still can influence adjacent regions of the protein to have higher entropy or threading energy. There was also no definite correlation between the sickle cell mutation in the β-subunit and presence of high threading energy at the site of mutation. The α-, β-, γ1-, ε-, and ζ-subunits all have roughly similar areas of high threading energy around residues 38–48, 58–73, 83–107, and 133–147, whereas the γ2- and δ-subunits have high threading energy around residues 38–45, 85–100, and 140–147. According to the analysis of modeled structures, the γ2- and δ-subunits have less prominent regions of high threading energy compared to the other subunits. The structural analysis of isoforms of γ-subunit generated rather unexpected results since these 2 subunits, being different at one position (the Gly136Ala polymorphism), were substantially different in their threading energy. The PTMs appear to be abundant and diverse in most of the subunits of hemoglobin, especially α- and β-chains. Reasonable correlation was found between the local predisposition of a protein for intrinsic disorder or increased conformational flexibility and position of various PTMs. There is an especially large number of PTMs surrounding the site of sickle cell mutation in the Hb β-subunit. It also seems that the subunits of adult hemoglobin have higher presence of different PTMs than the Hb chains found at the fetal and embryonic developmental stages. Finally, interactomes deduced by UniProt STRING internet tool and APID software helped us to find some possible correlations between sickle cell disease, intrinsic disorder, and other properties of each subunit of hemoglobin and their overall functions based on the presence of specific protein partners. Of all subunits, the α- and β-chains were the most promiscuous binders (they have 73 and 54 binding partners, respectively). Therefore, it is likely that the sickle cell mutation affects versatile and numerous functions of the Hb β-subunit. In fact, some of the

β-chain partners are responsible for making sure that this protomer binds to the α-subunits, others regulate transport of ammonia, prevent oxidative stress, regulate gene expression, promote or reduce transcription activity, produce megakaryocytes, and facilitate receptor-mediated endocytosis. Therefore, the sickle cell mutation could result not only in the lowered binding affinity to oxygen and the disrupted morphology of red blood cells, but also can change the gene expression, distort interaction of β-subunit with essential partners, and promote oxidative stress in erythrocytes. In other words, there is a possibility that one sickle cell mutation in the Hb β-chain can generate the domino effect leading to the numerous apparently unrelated dysfunctions.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

References

[1] Roseff SD. Sickle cell disease: a review. Immunohematology 2009; 25:67-74; PMID:19927623

[2] Angastiniotis M, Modell B, Englezos P, Boulyjenkov V. Prevention and control of haemoglobinopathies. Bull World Health Organ 1995; 73:375-86; PMID:7614670

[3] Serjeant GR. Sickle-cell disease. Lancet 1997; 350:725-30; PMID:9291916; http://dx.doi.org/10.1016/S0140-6736(97)07330-3

[4] Diallo D, Tchernia G. Sickle cell disease in Africa. Curr Opin Hematol 2002; 9:111-6; PMID:11844993; http://dx.doi.org/10.1097/00062752-200203000-00005

[5] Streetly A, Latinovic R, Hall K, Henthorn J. Implementation of universal newborn bloodspot screening for sickle cell disease and other clinically significant haemoglobinopathies in England: screening results for 2005–7. J Clin Pathol 2009; 62:26-30; PMID:19103854; http://dx.doi.org/10.1136/jcp.2008.058859

[6] Hassell KL. Population estimates of sickle cell disease in the US. Am J Prev Med 2010; 38:S512-21; PMID:20331952; http://dx.doi.org/10.1016/j.amepre.2009.12.022

[7] Pawliuk R, Westerman KA, Fabry ME, Payen E, Tighe R, Bouhassira EE, Acharya SA, Ellis J, London IM, Eaves CJ, et al. Correction of sickle cell disease in transgenic mouse models by gene therapy. Science 2001; 294:2368-71; PMID:11743206; http://dx.doi.org/10.1126/science.1065806

[8] Arends T, Lehmann H, Plowman D, Stathopoulou R. Haemoglobin North Shore-Caracas β 134 (H12) valine replaced by glutamic acid. FEBS Lett 1977; 80:261-5; PMID:891976; http://dx.doi.org/10.1016/0014-5793(77)80453-5

[9] Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998:473-84; PMID:9697205

[10] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999; 293:321-31; PMID:10550212; http://dx.doi.org/10.1006/jmbi.1999.3110

[11] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000; 41:415-27; PMID:11025552; http://dx.doi.org/10.1002/1097-0134(20001115)41:3%3c415::AID-PROT130%3e3.0.CO;2-7

[12] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al. Intrinsically disordered protein. J Mol Graph Model 2001; 19:26-59; PMID:11381529; http://dx.doi.org/10.1016/S1093-3263(00)00138-8

[13] Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002; 27:527-33; PMID:12368089; http://dx.doi.org/10.1016/S0968-0004(02)02169-2

[14] Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively disordered proteins. In: Buchner J, Kiefhaber T, eds. Handbook of Protein Folding. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA 2005:271-353

[15] Uversky VN, Dunker AK. Understanding protein non-folding. Biochim Biophys Acta 2010; 1804:1231-64; PMID:20117254; http://dx.doi.org/10.1016/j.bbapap.2010.01.017

[16] Uversky VN, Dunker AK. The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. F1000 Biology Reports 2013; 5:1; PMID:23361308; http://dx.doi.org/10.3410/B5-1

[17] Hemmings HC, Jr, Nairn AC, Aswad DW, Greengard P. DARPP-32, a dopamine- and adenosine 3′:5′-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus. J Neurosci 1984; 4:99-110; PMID:6319628

[18] Gast K, Damaschun H, Eckert K, Schulze-Forster K, Maurer HR, Muller-Frohne M, Zirwer D, Czarnecki J, Damaschun G. Prothymosin α: a biologically active protein with random coil conformation. Biochemistry 1995; 34:13211-8; PMID:7548085; http://dx.doi.org/10.1021/bi00040a037

[19] Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT, Jr. NACP, a protein implicated in Alzheimer disease and learning, is natively unfolded. Biochemistry 1996; 35:13709-15; PMID:8901511; http://dx.doi.org/10.1021/bi961799n

[20] Williams RM, Obradovi Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker AK. The protein non-folding problem: amino acid determinants

of intrinsic order and disorder. Pac Symp Biocomput 2001:89-100; PMID:11262981

[21] Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001; 42:38-48; PMID:11093259; http://dx.doi.org/10.1002/1097-0134(20010101)42:1%3c38::AID-PROT50%3e3.0.CO;2-3

[22] Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007; 92:1439-56; PMID:17158572; http://dx.doi.org/10.1529/biophysj.106.094045

[23] Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics 2007; 8:211; PMID:17578581; http://dx.doi.org/10.1186/1471-2105-8-211

[24] Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al. UniProt: the Universal protein knowledgebase. Nucleic Acids Res 2004; 32:D115-9; PMID:14681372; http://dx.doi.org/10.1093/nar/gkh131

[25] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011; 7:539; PMID:21988835; http://dx.doi.org/10.1038/msb.2011.75

[26] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010; 1804:996-1010; PMID:20100603; http://dx.doi.org/10.1016/j.bbapap.2010.01.011

[27] Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 2005; 3:35-60; PMID:15751111; http://dx.doi.org/10.1142/S0219720005000886

[28] Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2005; 61(Suppl 7):176-82; PMID:16187360; http://dx.doi.org/10.1002/prot.20735

[29] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005; 21:3433-4; PMID:15955779; http://dx.doi.org/10.1093/bioinformatics/bti541

[30] Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SC. Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics 2015; 31:201-8; PMID:25246432; http://dx.doi.org/10.1093/bioinformatics/btu625

[31] Fan X, Kurgan L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J Biomol Struct Dyn 2014; 32:448-64; PMID:23534882; http://dx.doi.org/10.1080/07391102.2013.775969

[32] Peng Z, Kurgan L. On the complementarity of the consensus-based disorder prediction. Pac Symp Biocomput 2012:176-87; PMID:22174273

[33] Di Domenico T, Walsh I, Martin AJ, Tosatto SC. MobiDB: a comprehensive database of intrinsic protein disorder annotations. Bioinformatics 2012; 28:2080-1; PMID:22661649; http://dx.doi.org/10.1093/bioinformatics/bts327

[34] Potenza E, Domenico TD, Walsh I, Tosatto SC. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. Nucleic Acids Res 2015; 43:D315-20; PMID:25361972

[35] Walsh I, Martin AJ, Di Domenico T, Tosatto SC. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 2012; 28:503-9; PMID:22190692; http://dx.doi.org/10.1093/bioinformatics/btr682

[36] Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure 2003; 11:1453-9; PMID:14604535; http://dx.doi.org/10.1016/j.str.2003.10.002

[37] Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005; 21:3369-76; PMID:15947016; http://dx.doi.org/10.1093/bioinformatics/bti534

[38] Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. Bmc Bioinformatics 2006; 7:208; http://dx.doi.org/10.1186/1471-2105-7-208

[39] Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Research 2003; 31:3701-8; PMID:12824398; http://dx.doi.org/10.1093/nar/gkg519

[40] Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, et al. DisProt: the database of disordered proteins. Nucleic Acids Res 2007; 35:D786-93; PMID:17145717; http://dx.doi.org/10.1093/nar/gkl893

[41] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acids Res 2014; 42:D222-30; PMID:24288371; http://dx.doi.org/10.1093/nar/gkt1223

[42] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000; 28:235-42; PMID:10592235; http://dx.doi.org/10.1093/nar/28.1.235

[43] Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, et al. D(2)P(2): database of disordered protein predictions. Nucleic Acids Res 2013; 41:D508-16; PMID:23203878; http://dx.doi.org/10.1093/nar/gks1226

[44] Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 2007; 35:W460-4; PMID:17567614; http://dx.doi.org/10.1093/nar/gkm363

[45] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39:D561-8; PMID:21045058; http://dx.doi.org/10.1093/nar/gkq973

[46] Alonso-Lopez D, Gutierrez MA, Lopes KP, Prieto C, Santamaria R, De Las Rivas J. APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. Nucleic Acids Res 2016; 44(W1):W529-35; PMID:27131791

[47] Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res 2015; 43:D470-8; PMID:25428363; http://dx.doi.org/10.1093/nar/gku1204

[48] Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. Nucleic Acids Res 2004; 32:D449-51; PMID:14681454; http://dx.doi.org/10.1093/nar/gkh086

[49] Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al. Human Protein Reference Database–2009 update. Nucleic Acids Res 2009; 37: D767-72; PMID:18988627; http://dx.doi.org/10.1093/nar/gkn892

[50] Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al. The IntAct molecular interaction database in 2012. Nucleic Acids Res 2012; 40:D841-6; PMID:22121220; http://dx.doi.org/10.1093/nar/gkr1088

[51] Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 2012; 40:D857-61; PMID:22096227; http://dx.doi.org/10.1093/nar/gkr930

[52] Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 2015; 162:425-40; PMID:26186194; http://dx.doi.org/10.1016/j.cell.2015.06.043

[53] Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 2015; 43:D345-56; PMID:25428375; http://dx.doi.org/10.1093/nar/gku1214

[54] Manning LR, Russell JE, Padovan JC, Chait BT, Popowicz A, Manning RS, Manning JM. Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. Protein Sci 2007; 16:1641-58; PMID:17656582; http://dx.doi.org/10.1110/ps.072891007

[55] Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting Protein Disorder for N-, C-, and internal regions. Genome Informatics Workshop Genome Informatics 1999; 10:30-40; PMID:11072340

[56] Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 2006; 7:208; PMID:16618368; http://dx.doi.org/10.1186/1471-2105-7-208

[57] Haller JO, Berdon WE, Franke H. Sickle cell anemia: the legacy of the patient (Walter Clement Noel), the interne (Ernest Irons), and the attending physician (James Herrick) and the facts of its discovery. Pediatr Radiol 2001; 31:889-90; PMID:11727028; http://dx.doi.org/10.1007/s002470100014

[58] Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: A Plausible Model. J Mol Biol 1965; 12:88-118; PMID:14343300; http://dx.doi.org/10.1016/S0022-2836(65)80285-6

[59] Perutz MF. Stereochemistry of cooperative effects in haemoglobin. Nature 1970; 228:726-39; PMID:5528785; http://dx.doi.org/10.1038/228726a0

[60] Fan JS, Zheng Y, Choy WY, Simplaceanu V, Ho NT, Ho C, Yang D. Solution structure and dynamics of human hemoglobin in the carbonmonoxy form. Biochemistry 2013; 52:5809-20; PMID:23901897; http://dx.doi.org/10.1021/bi4005683

[61] Pettigrew DW, Turner BW, Ackers GK. Subunit assembly and interactions in hormal and abnormal human hemoglobins. In: Sigler PB, ed. The Molecular Basis of Mutant Hemoglobin Dysfunction. New York, Amsterdam, Oxford: Elsevier/North-Holland, Inc., 1981:191-8

[62] Silva MM, Rogers PH, Arnone A. A third quaternary structure of human hemoglobin A at 1.7-A resolution. J Biol Chem 1992; 267:17248-56; PMID:1512262

[63] Safo MK, Abraham DJ. The enigma of the liganded hemoglobin end state: a novel quaternary structure of human carbonmonoxy hemoglobin. Biochemistry 2005; 44:8347-59; PMID:15938624; http://dx.doi.org/10.1021/bi050412q

[64] Chu X, Wang J. Specificity and affinity quantification of flexible recognition from underlying energy landscape topography. PLoS Comput Biol 2014; 10:e1003782; PMID:25144525; http://dx.doi.org/10.1371/journal.pcbi.1003782

[65] Chu X, Gan L, Wang E, Wang J. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. Proc Natl Acad Sci U S A 2013; 110: E2342-51; PMID:23754431; http://dx.doi.org/10.1073/pnas.1220699110

[66] Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004; 32:1037-49; PMID:14960716; http://dx.doi.org/10.1093/nar/gkh253

[67] Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, Radivojac P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci 2014; 23:1077-93; PMID:24888500; http://dx.doi.org/10.1002/pro.2494

[68] Harano T, Harano K, Kushida Y, Ueda S, Kawakami H. Hb A2-Niigata [delta 1(NA1)Val—-Ala]: a new delta

chain variant found in the Japanese population. Hemoglobin 1991; 15:335-9; http://dx.doi.org/10.3109/03630269109027889

[69] Ilesanmi OO. Pathological basis of symptoms and crises in sickle cell disorder: implications for counseling and psychotherapy. Hematol Rep 2010; 2:e2; PMID:22184515; http://dx.doi.org/10.4081/hr.2010.e2

[70] Bunn HF. Pathogenesis and treatment of sickle cell disease. N Engl J Med 1997; 337:762-9; PMID:9287233; http://dx.doi.org/10.1056/NEJM199709113371107