ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Hemoglobin complex analysis: structural properties and interactions within the human proteome

Giorgia Del Missier[*,1]
[*]Department of Pharmacy and Biotechnology, University of Bologna

ABSTRACT Hemoglobin plays a central role for human and mammalian life thanks to its ability to carry oxygen to the tissues. This occurs through the binding of O_2 to the four heme groups of the quaternary protein. Using the information stored in the PDB and DSSP files of its structure, a computational analysis of the tetramer was performed in order to identify the key components that contribute to the stabilization of the complex, in particular salt bridges and hotspots. Moreover, bioinformatic tools were used in order to better understand the interactions occurring between the protein monomers and the heme groups associated with them. Finally, an analysis of the connectivity of the alpha and beta subunits of hemoglobin within the human proteome was carried out, obtaining insights on the implication of these proteins in different biological pathways.

KEYWORDS
HEMOGLOBIN
SALT BRIDGES
HOTSPOTS
PROTEOME

INTRODUCTION

The transport of oxygen from the environment to the cells and the transport of the metabolically produced carbon dioxide in the opposite direction is essential for vertebrate life. Hemoglobin, abbreviated Hb (or Hgb), is a protein present in red blood cells of almost all vertebrates which is able to carry oxygen from the lungs to the tissues. There, it releases O_2, which is used as a terminal electron acceptor in the production of ATP, thus permitting oxidative phosphorylation and aerobic respiration. ATP produced is then used as energy to power all the functions the organism needs during metabolism.

Due to its crucial role for life, Hgb has become one of the most intensively studied and characterized proteins, which resulted in a deep understanding of its structure and function.

Adult human hemoglobin is a tetrameric globular protein (Figure S1), consisting of two α and two β polypeptide chains, composed respectively of 141 and 146 amino acids. Each chain is made up of alternating α-helical segments (labeled A to H from the N terminus) and non-helical ones (named by the letters of the adjacent helices). The final N- and C- termini are also non-helical (Figure 1). α and β subunits are very similar, differing just by an additional helix present only in the latter. The final conformation has a roughly tetrahedral shape (Jensen *et al.* 1998).

To each of the four subunits, a non-protein prosthetic heme group is associated. A heme group consists of an iron (Fe) ion held in a heterocyclic ring, called porphyrin, consisting of four pyrrole molecules. The heme group is found within the folded protein, buried inside a hydrophobic pocket between the E and F helices of each chain. Different residues, the most important being a proximal histidine (F8 His), contribute to the binding of the iron atom to the globular protein (Jensen *et al.* 1998).

The fine-tuning of the assembly process of this multisubunit protein is a critical post-translational event: in erythroblasts, there exist a perfect balanced synthesis of the α and β subunits. In the most probable assembly process, $\alpha\beta$ dimers are the first to form, this process facilitated by electrostatic attractions between the positively charged alpha subunit and the negatively charged beta subunit (Mrabet *et al.* 1985). At comparable rates, also porphyrin rings are synthesized; the iron atom successively combines with it to form the final heme group. The association of the heme and the dimer occurs immediately after; finally, the $\alpha_2\beta_2$ tetramer is assembled (Felicetti *et al.* 1966).

The iron atom is the site where the ligand binds. Hence, each hemoglobin can carry four oxygen atoms: in this saturated form, the protein is referred to as oxyhemoglobin, as opposed to the desaturated form, deoxyhemoglobin.
X-ray crystallography revealed that the quaternary structure of the unliganded Hb differs from the one bound to oxygen (Jensen *et al.* 1998). In fact, hemoglobin can exist in equilibrium between two states: the taut (or tense) form, denoted T, and the relaxed one, R (Figure 2). The former in characterized by a low O_2 affinity and is thermodynamically more stable. The latter is able to bind oxygen: upon ligand binding, a slight conformational change in the Hgb

Manuscript compiled: Friday 14th February, 2020

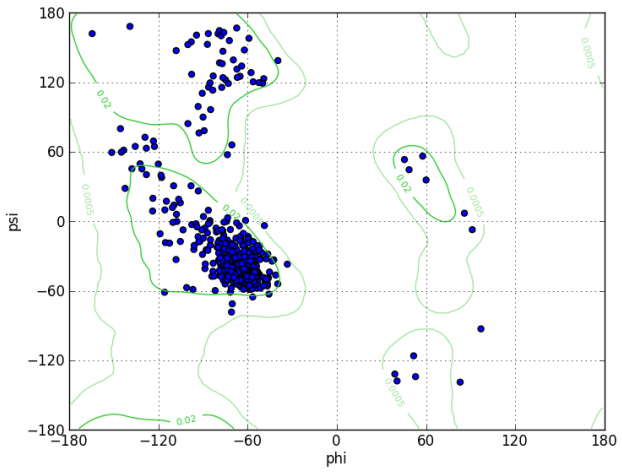

Figure 1 Ramachandran plot of the hemoglobin complex, showing that the majority of residues is found in an α-helix secondary structure.

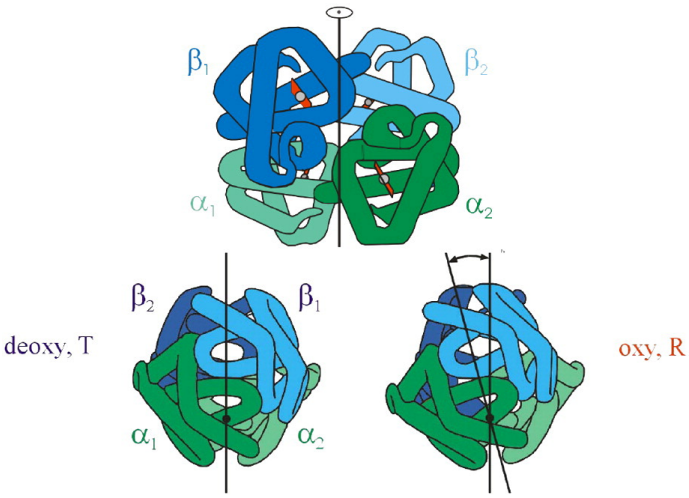

Figure 2 The two different conformations of oxy- and deoxy-hemoglobin.

structure occurs, encouraging O_2 to bind to the three remaining heme units (allosteric model).

It is thus worth to further investigate the mechanisms that contribute to the stabilization and the maintenance of the quaternary complex and the extent to which the different chains interact in order to better grasp the possible events that favor ligand binding and are responsible for the positive cooperative process that follows.

Moreover, this study highlights the wide range of connections occurring not only between the two hemoglobin subunits but also with a vast variety of other proteins involved in several biological pathways.

MATERIALS AND METHODS

Several bioinformatic tools were employed during this analysis of the hemoglobin protein.

In order to examine the structural data, the Python programming language (version 3.7.2) was used. Python scripts, available in the set of supplementary materials, were produced in order to evaluate the extent of interaction between the subunits, their surface of contact and their involvement in the different protein networks. The Numpy and NetworkX libraries were also exploited during some steps of the analysis.

Data handling and programs launching was always performed through the use of Unix commands from the terminal window.

Also, the DSSP program (version DsspCMBI-December-2001) was used in order to assign secondary structures, geometrical features and solvent exposure of the amino acids of the protein from their atomic coordinates.

Finally, validation was performed using the visualization software UCSF Chimera (version 1.14): 3D images were produced to better describe the putative interactions occurring between the protein residues and chains.

PDB parsing

The analysis starts from the description of the structure of the hemoglobin complex obtained from the Protein Data Bank (PDB). PDB is a large, public database storing 3D structural data of biological molecules, such as proteins and nucleic acids, usually obtained through experimentally techniques, namely X-ray crystallography, NMR spectroscopy and cryo-electron microscopy. The .pdb file format provides information on the coordinates of the atoms that are part of the protein as well as the ones of the hetero-atoms.

The PDB structure for the oxygenated T state adult human hemoglobin was retrieved from the PDB website using the identification code **1GZX**. The structure was obtained through X-ray crystallography at a resolution of 2.1 Angstroms.

The hemoglobin molecule in the tense state would be normally deoxygenated; however, during crystallization a protocol was developed in order to oxygenate the molecule (Paoli *et al.* 1996).

The PDB hemoglobin file discriminates between the atoms that are part of the polypeptide and the ones of the prosthetic heme group by labelling them respectively as "ATOM" and "HETATM". Also, the residues were subdivided into four distinct groups depending on the chain they belonged to: A and C corresponded to the two α subunits, B and D to the β subunits. The file was parsed by means of the python scripts and general information including chain ownership, residue type and number and atomic coordinates were stored. Two different dictionary data structures were created for the protein residues and for the heteroatoms. Water molecules were excluded from the parsing.

The analysis of the interacting residues was performed both between all the distinct monomers and between the latter and their corresponding heme and oxygen. Pairwise distances in the three dimensional space were computed using the formula for the Euclidean distance:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \qquad (1)$$

where the vectors (x_1, y_1, z_1) *and* (x_2, y_2, z_2) *represent the spatial coordinates of the single points.*

The accepted donor-acceptor distance for molecular interaction is around 3Å. However, PDB does not include the representation of hydrogen atoms: for this reason, residues separated by less than 3.5Å were considered significant.

The putative interacting atoms were reported as output.

DSSP parsing

In order to identify the surface of interaction between the monomers of the hemoglobin complex, the .dssp file for-

mat was used. This file extension derives from the name of the algorithm designed by W. Kabsch and C. Sander (https://github.com/cmbi/hssp/releases) which generates it. The DSSP (Define Secondary Structure of Proteins) program identifies intra-backbone hydrogen bonds of the protein using a simple electrostatic equation. Based on this result, the secondary structure of each atom is assigned.
For the aim of this analysis, the crucial piece of information stored in the these files was the one regarding the accessible surface area (ASA), i.e. the surface area of a biomolecule accessible to a solvent. The DSSP algorithm is one of the most commonly used programs for generating such kind of data.

To assess the extent of solvent accessibility for each chain in the tetramer, the DSSP files for the full hemoglobin , for all possible trimers lacking one subunit and for each monomer were produced. These tasks were performed using basic Unix commands.
The differences in ASA values in the three separate complexes can be then evaluated, since these will vary according to whether the amino acids in that specific conformation are interacting with other residues or not.

To run the analysis on these files, a new parsing python script was generated.
To compute the surface of interaction between each of the two monomers, the total ASA is found both for the tetramer and for each trimer: their difference is used to evaluate loss of solvent accessibility.
To determine residues that are likely to interact, the relative accessible surface area (RASA) for each amino acid was computed, both in the quaternary structure and in the single chain. The RASA of a protein residue is a measure of the residue solvent exposure, normalized by its maximum possible ASA. This value is usually obtained from tripeptides composed by the amino acid of interest bound to two molecules of glycine. The maximum ASA were retrieved from AAIndex, a database storing various physicochemical and biochemical properties of amino acids and pairs of amino acids. Only residues that lost 10% of relative accessible area in the complex with respect to the single monomer were taken into consideration.

Protein-protein interaction network analysis

The proteomic subnetwork in which the α and β subunits of human hemoglobin was investigated included only the latter ones and the proteins directly interacting with them. This analysis was perfomed downsizing a mitab file downloaded from the IntAct database containing information on all known pairwise interactions existing between proteins.
Also in this case, a python script was produced in order to parse the file, filter it for redundancy and take into account only proteins of human origin.
The connectivity between the proteins was evaluated through the generation of an undirected network. Finally, different properties of the subnet like betweenness centrality, node degree and clustering coefficient were produced for the alpha and beta subunits and for the proteins directly interacting with them.

RESULTS

Interactions between monomers and hetero groups

Hemoglobin has a quaternary structure, with each of the four monomers linked to a prosthetic group which is bound, as previously mentioned, by a hydrophobic pocket. In order to further investigate this interaction, residues in each chain found within 3.5Å from the heme and the oxygen atoms were computed (Table 1).

The obtained data highlights a symmetry existing between the different subunits, with similar residues found close to the heme group – in particular, His, Leu and Asn. Chain A is the one making the highest number interactions, with more than 10 residues contributing to the stabilization. Only two residues for each monomer – i.e. His and Val – were found in close proximity to the oxygen atom.

From the results it also emerges the crucial role performed by two histidines per monomer in order to anchor the hetero group: the first (His87.A, His235.B, His487.C and His635.D) contributes to the stabilization by interacting with the iron through a nitrogen atom; the second (His58.A, His206.B, His458.C, His606.D) is found next to the bound oxygen (Figure 3).

Moreover, the hydrophobic residues forming the pocket to which the hetero group is bound have been investigated in chain A – similar results are expected for the other subunits. Apolar side chains of the Phe, Val and Leu residues seem to form a ring above

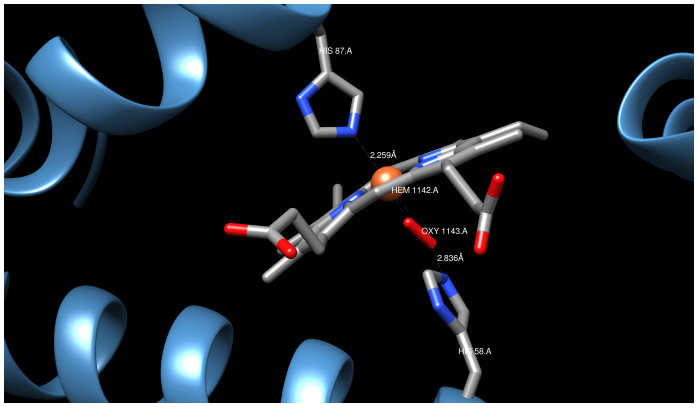

Figure 3 His87.A and His58.A interaction with the heme group and the oxygen atom. This is one of the main factors contributing to stabilization.

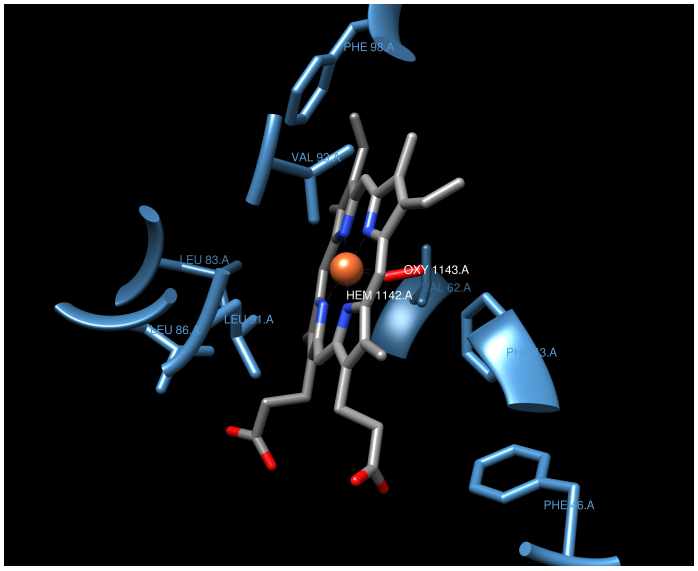

Figure 4 The hydrophobic pocket around the hetero group that anchors it to the complex. In this picture, chain A; similar results are obtained for the other monomers.

Table 1 Heme - Monomer interactions

For each chain, the most significant amino acids making contact with the oxygenated heme groups (and the respective atoms for both residues) have been reported. Full tables on Supplementary materials.

Chain	Residue	Hetero	Atoms ($\leq$ 3.5Å)
A	HIS58	HEM1142	CE1-CHA, CE1-C1A, CE1-C4D
A	HIS87	HEM1142	CE1-C4A, NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
A	HIS58	OXY1143	NE2-O1, NE2-O2
A	VAL62	OXY1143	CG2-O2
B	HIS206	HEM1290	CE1-CHA, CE1-C1A, NE2-C4D
B	HIS235	HEM1290	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
B	HIS206	OXY1291	NE2-O1, NE2-O2
B	VAL210	OXY1291	CG2-O1, CG2-O2
C	HIS487	HEM1542	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
C	HIS458	OXY1543	NE2-O1, NE2-O2
C	VAL462	OXY1543	CG2-O2
D	LEU649	HEM1690	CD2-CHC, CD2-CBB, CD2-CMC
D	HIS635	HEM1690	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
D	HIS606	OXY1691	NE2-O1, NE2-O2
D	VAL610	OXY1691	CG2-O1, CG2-O2

and below the prostethic group, stabilizing the interaction (Figure 4).

Monomer - monomer interaction

The same procedure followed in the previous section was applied in order to find pairwise monomer non-covalent bonds. As before, 3.5Å was set as the threshold of significance for the distances between putative interacting residues.

The residues found at closest distance for each pair of monomers are reported in Table 2.

The obtained results emphasize the symmetry existing between the two dimers that make up the final complex: both A and C interact with their respective β subunits (chains B and D) by means of the same residue types, namely Arg, Pro and Phe. Chain A also makes contact with residues of the other α subunit and with chain D; no significative interaction occurs between the two beta monomers.

Monomers surface of interaction

The following step of the analysis was aimed at the quantification of the surface of interaction between the single chains through the use of the DSSP files. In particular, the amount of surface between each pair of monomers was found as the difference between the total ASA of the complex and the one in the trimer lacking the subunit under analysis. Data is reported in Table 3.

Ideally, the surface of interaction between the two monomers should be found to be symmetric; however, some degree of deviation was observed, probably due to molecular measurement made by the DSSP algorithm.

A general trend can be identified from the results:

- The interface between the subunits forming the $\alpha\beta$ dimer (A-B and C-D) is very large (with a value averaging around 850Å), suggesting the high stability present between these two subunits.
- The high degree of interaction existing between the two dimers to form the final tetramer complex is testified by the large extent of SA existing between the A-D and B-C chains.
- The two alpha chains share a small portion of contact (around 240Å), while the surface of interaction between the two beta chains is almost null (around 24Å).

These results are consistent with the ones found in the previous section; furthermore, they seem to confirm the most probable process of assembly as described in the introduction.

Table 3 Surface of interaction between all pairs of monomers

Chain1	Chain2	SA (Ch1-Ch2,in Å^2)	SA(Ch2-Ch1)
A	B	862.0	885.0
A	C	244.0	240.0
A	D	696.0	664.0
B	C	665.0	680.0
B	D	23.0	24.0
C	D	813.0	833.0

Residues at the interacting surface

To determine the residues that are involved in the formation of the $\alpha\beta$ dimer and the ones that contribute to the stabilization of the final quaternary complex, a similar analysis was performed. In this

Table 2 Monomer - Monomer interactions

Pairwise interactions between each monomer. Full tables on Supplementary materials.

Chain1	Res1	Chain2	Res2	Atoms (≤ 3.5Å)
A	ARG31	B	GLN270	CB-OE1, CD-OE1, NH1-OE1
A	ARG31	B	PHE265	CZ-O, NH1-O, NH2-O
A	PRO114	B	HIS259	CA-NE2, C-NE2, O-NE2
A	PHE117	B	ARG173	O-NH2, CD2-NH2, CE2-NH2
A	ARG141	C	LYS527	C-NZ, O-NZ, OXT-NZ
A	ARG141	C	ASP526	CZ-OD2, NH1-OD2, NH2-OD2
A	ARG92	D	ARG583	O-CD, CB-CD, CZ-CB, NH1-CA
A	ARG141	D	VAL577	NE-O, CZ-O, NH1-CG1
B	ASP242	C	TYR442	CB-OH, CG-OH, OD1-OH
C	ARG431	D	GLN670	CB-OE1, CG-OE1, CD-OE1, CZ-OE1, NH1-OE1
C	ARG431	D	PHE665	CZ-O, NH1-O, NH2-O
C	PRO514	D	HIS659	CA-NE2, C-NE2, O-NE2
C	PHE517	D	ARG573	O-NH2, CD2-NH2, CE2-NH2
C	HIS522	D	ARG573	CB-NH2, CG-NH2, ND1-NH2

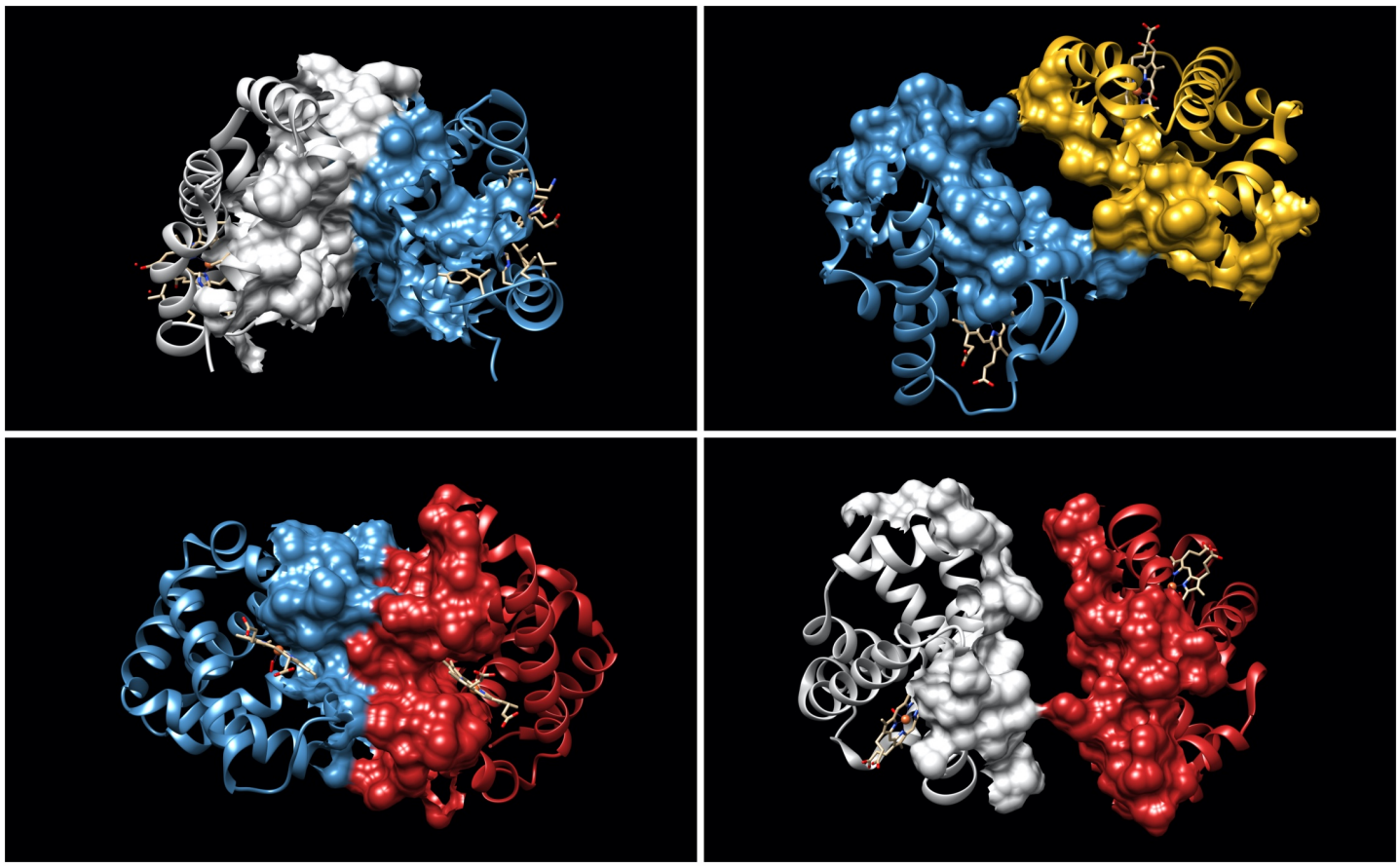

Figure 5 Chain A (blue) interacting respectively with chain B (gray), chain C (gold) and chain D (red). In the last picture it can be seen how the there is almost no interaction between chains B and D.

Table 4 Residues with significant loss of RASA
Full tables on Supplementary materials.

Chain	Residue	RSA(M)	RSA(C)	RSA(M) - RSA(C ($\leq$ 3.5Å)
A	SER35	0.687	0.1391	0.5478
A	THR41	0.7571	0.0571	0.7
A	ARG92	0.7467	0.1911	0.5556
A	ASP94	0.58	0.0533	0.5267
A	PRO119	0.7655	0.1448	0.6207
A	ARG141	1	0.3422	0.6578
B	VAL177	0.5548	0.0065	0.5484
B	TRP180	0.6235	0.0745	0.549
B	ASP242	0.6267	0.0133	0.6133
B	CYS255	0.6148	0.0222	0.5926
B	PRO267	0.6966	0.1103	0.5862
C	SER435	0.6522	0.1304	0.5217
C	THR441	0.7286	0.0643	0.6643
C	ARG492	0.7333	0.2044	0.5289
C	PRO519	0.7241	0.1103	0.6138
C	ARG541	1	0.3556	0.6444
D	VAL577	0.529	0.0	0.529
D	TRP580	0.6196	0.0745	0.5451
D	ASP642	0.64	0.0133	0.6267
D	CYS655	0.6519	0.0222	0.6296
D	PRO667	0.6621	0.1103	0.5517

case, the evaluation was carried out by comparing the percent of relative accessible surface area in the tetramer and in the monomer. Only residues that showed a difference higher than the 10% were taken into account.
Some of the residues, in particular the ones found at the N- and C- termini, showed a value of lost RASA higher than 1 – i.e. they exceeded the maximum ASA retrieved from AAIndex: this is mostly due to their position on the chain and maybe to some measurement bias. In this cases, 1 was taken by default as RASA value.

For each chain, the most significant residues (lost RASA > 50%) found in this analysis are reported in Table 4.

Also in this case, the symmetry that characterizes the complex appears evident: the two alpha chains (monomers A and C) show the highest RASA percentage loss in the same residue types and the same occurs for the beta monomers (B and D).

Furthermore, when considering all the significant AAs, it should be noted that these are always separated by two or three residues. Each monomer is mainly formed by repeated alpha helices, which have a turn of around 3.6 residues: hence, the identified amino acids are all on the same face of the helix.
This is consistent with the hypothesis that these residues lose a high percentage of RASA as a result of the interaction they make with close to monomers (Figure 6).

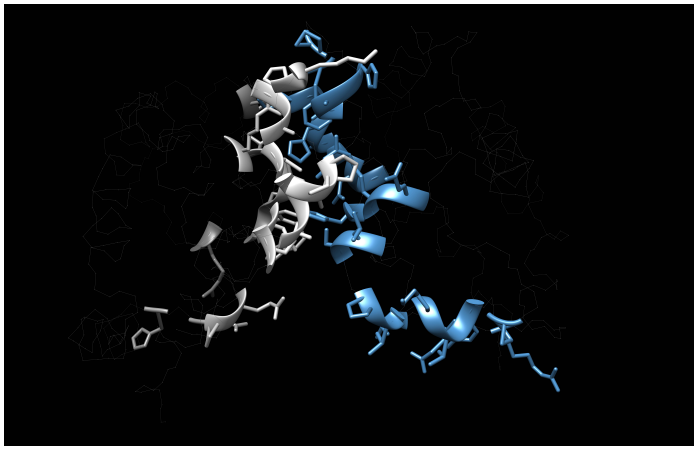

Figure 6 Residues of chain A and B which lose more than 10% of RASA when found in the tetramer complex. As it can be seen in this picture, all of them are found on the same face of the α-helix, which is also the one making contact with the helix in the other component of the dimer.

Table 5 Putative salt bridges identified by the computational analysis

Chain 1	Residue 1	Chain 2	Residue 2	Distance (Å)
A	GLU23	B	LYS263	3.929
A	ASP126	C	ARG541	2.440
A	LYS127	C	ARG541	2.775
A	ARG141	C	ASP526	2.814
A	ARG141	C	LYS527	2.387
A	ARG92	D	GLU586	2.653

Salt bridges identification

By combining the results obtained in the previous sections, some relevant non-covalent bonds contributing to the process of stabilization can be identified. In particular, salt bridges might play a crucial role not only in the creation of the dimer, but also in the formation of a well-anchored tetramer complex.

Salt bridges are a combination of hydrogen and ionic bonds. In fact, they are defined as non-covalent bonds between oppositely charged residues that are close enough to be electrostatically attracted. Hence, they are mainly composed of negative charges from Asp, Glu and the C-terminal carboxylate groups and of positively charged His, Lys, Arg functional groups and N-terminal amino groups. Since the side chain charge depends on pH, these bonds are said to be pH-dependent.
Together with hydrophobic effect, van der Waals interaction and hydrogen bonds, salt bridges are one of the main contributors to protein stability (Bosshard *et al.* 2004).

To identify salt bridges present in hemoglobin, the list of residues at distance less than 3.5Å was compared with the one comprising amino acids with a significative loss of RASA. The intersection between these two groups was screened for residues of opposite charge. Moreover, special care was applied on the residues at the N- and C- termini.

As can be seen in Table 5, chain A is the only monomer to interact through salt bridges with all the other three polypeptides.
In particular, the most relevant finding is the interaction made between the two α chains. Stabilization is achieved through four salt bridges: the first is made between the positively charged side chain of Arg141.A and the negatively charged functional group of Asp526.C; the second occurs between the C-terminal carboxylate group of Arg141.A and Lys527.C.
Symmetrically, the C-terminal residue of chain C (Arg541) is able to form two salt bridges: one through it side chain with Asp126.A and the other through its free carboxyl group with Lys127.A. Both carbon terminal residues, hence, insert their C-tail into the opposite monomer and this results in the stabilization of the complex (Figure 7, A).

Moreover, also Arg92.A and Glu586.D are in perfect orientation to form a salt bridge (Figure 7, B).
Finally, also residues Glu23.A and Lys263.B interact, probably in order to stabilize the dimer: in this case, the distance is a bit higher than the threshold of 3.5Å but this can be explained by the absence of the hydrogen atoms in the PBD file (Figure 7, C).

To assess the existence of other salt bridges not identified by the python script because formed by residues at the N- or C- termini, manual validation through Chimera was performed.
From this analysis, it emerged that chain A and D form a salt bridge through Lys40.A and the carboxyl group of His689.D and chain B and C interact by means of Lys440.C and His289.B. All the other residues either were not in the right orientation or they were too far from one another.

The existence of salt bridges in the hemoglobin tertiary structure has long been debated in the scientific community.
Max F. Perutz, the first to determine the structure of the hemoglobin through X-ray crystallography in 1959, was also able to identify a number of salt bridges which are present in the deoxyhemoglobin but are absent when the ligand binds.
According to this hypothesis, the C-termini residues of each chain (i.e. Arg141.A, Arg541.C and His289.B, His689.D) form additional salt bridges to better stabilize the quaternary complex in the T state (Bettati *et al.* 1998).
The hypothesis then suggests that the binding of O_2 to the heme group induces a movement of the iron atom and its linked His. This transition also causes a rotation of the $\alpha\beta$ dimer of about 15 degrees (Figure 2), which also shifts the contacts between the different subunits. The model than postulates that breakage of these salt bridges are able to increase the oxygen affinity of the taut structure, hence producing a cooperative effect in the binding of the ligand (Perutz 1990).

The results obtained in this study are perfectly consistent with this hypothesis: the structure under study represents an adult hemoglobin in the T state and salt bridges occurring between chains belonging to opposite dimers have been detected by the computational approach used in this analysis.

Hotspots identification

Another important piece of information that can be inferred from the analysis of the interacting amino acids in the complex is the identification of the hotspots.

The sites through which proteins associate are usually characterized by adequate electrostatic complementarity and shape. In the context of protein-protein interaction, it has been shown that the energy distribution is not uniform and some residues are more important than others for stabilization. These amino acids are usually identified through alanine scanning, which consists in replacing each residue in the interface by alanine and computing the difference in binding free energy between the wild type and the mutant. Alanine is used since it is a relatively inert amino acid, with a simple methyl functional group. Hotspots are then defined as residues whose substitution leads to a significant drop in binding free energy (usually $\geq$ 2kcal/mol) (Keskin *et al.* 2007).
The composition of hotspots is not random: usually PPI are hydrophobic and some residues, such as tryptophan (21%), arginine (13.3%) and tyrosine (12.3%), have been found to be much frequent than the others. In particular, tryptophan is a large, hydrophobic amino acid with an aromatic group able to interact with other

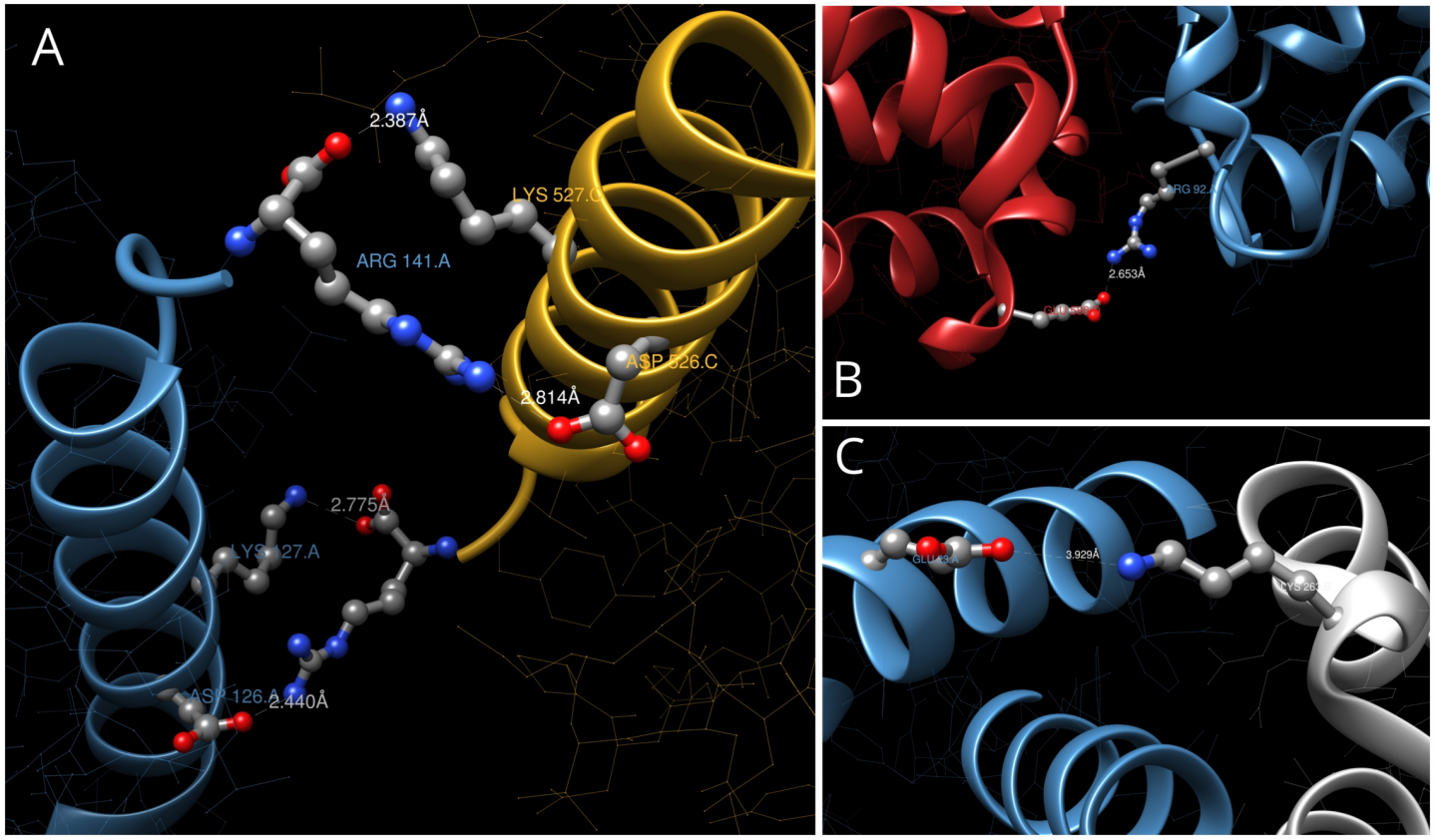

Figure 7 Visualization of the computationally predicted salt bridges through the Chimera software.
(A) The two α subunits stabilize their interaction through two pairs of salt bridges, formed respectively by the charged functional group and the free carboxyl group of their two C-terminal residues. (B) The salt bridge existing between chain A (blue) and chain D (red). (C) Dimer anchoring is also due to the presence of salt bridges, as exemplified by the one existing between Glu23.A and Lys263.B.

molecules. When Trp is mutated to alanine, the high difference in size generates a hole in the protein, causing a big destabilization in the complex (Morrow and Zhang 2013).
The identification of hotspots is important to spot potential active sites, mainly useful for drug target. However, experimental techniques used to identify them are still quite slow and expensive and computational approaches are usually preferred.

In this case, putative hotspots have been defined as residues that lost a significative percentage of RASA and were found at a maximum distance of 4Å from AAs in the other interacting chains. Moreover, both the residues under analysis had to have an apolar side chain (Table 6).

From these results, it emerges that most of the hotspots are located within the interacting surface of the two dimers.
This data has been successively validated through the use of Chimera. As it can be seen in Figure S2, which illustrates the AB dimer, on top Pro119.A and Ala123.A are shown to interact with Val176.B, Val177.B and Met198.B. In the middle, Val107.A and Ala111.A and Val254.B, Ala258.B and Gly262.B are located next to each other. Finally, in the bottom, Leu34.A interacts with close by residues in chain B – namely Pro267, Pro268 and Ala271. Similar results are obtained when considering the CD dimer.

The final complex, instead, seems not to be stabilized by hotspots, a part from a single residue in chain B (Trp180.B) which is predicted to interact with Pro495 of chain C.

Hemoglobin interactions in the proteome network

Finally, the level of the connectivity of the α and β subunits within the human proteome was assessed.

Network properties were computed by parsing a file from the IntAct database containing information on all pairwise interactions between proteins and selecting only the ones of human origin.

The network was firstly downsized to a subnet consisting only of proteins with a maximum of two degrees of separation from the two hemoglobin subunits alpha and beta (identified, respectively, by their UniProt ID: **P69905** and **P68871**) (Figure 8).
For both, the degree, clustering coefficient and betweenness centrality were calculated using NetworkX standard methods.

In graph theory, the topology of the network can in fact be defined by some important properties.
In particular the level of connectivity among the nodes – which, in this case, represent the proteins – depends on the number of the edges – i.e. the interactions between them.

The degree is hence defined as the number of connections a node makes with the others present in the network.

Another important property of a graph is the so-called clustering coefficient (or transitivity). This is a measure of the level to which nodes tend to cluster together: in practice, it compares the observed level of connectivity of a node with the total possible theoretical number of edges in it.

Table 6 Putative hotspots identified by computational analysis

Residues highlighted in bold are the ones found also using as maximum distance 3.5Å.

Chain 1	Residue 1	Chain 2	Residue 2 (Å)
A	**LEU34**	**B**	**ALA271**
A	LEU34	B	PRO267
A	LEU34	B	PRO268
A	VAL107	B	ALA258
A	VAL107	B	VAL254
A	ALA111	B	ALA258
A	**ALA111**	**B**	**GLY262**
A	PRO119	B	VAL176
A	PRO119	B	MET198
A	ALA123	B	VAL177
B	TRP180	C	PRO495
C	LEU434	D	ALA671
C	LEU434	D	PRO667
C	VAL507	D	ALA658
C	**ALA511**	**D**	**ALA658**
C	**ALA511**	**D**	**GLY662**
C	PRO519	D	VAL576
C	ALA520	D	PRO594
C	ALA523	D	VAL576
C	ALA523	D	VAL577

It is defined as:

$$c_i = \frac{2e_i}{k_i(k_i - 1)} \tag{2}$$

where the k_i is the number of nodes connected to node i and e_i the number of edges between the k_i nodes.

The last property taken into consideration is the betweenness centrality, which is a general measure of centrality. For every pair of nodes, there exist at least one shortest path between them. This property then defines, for each vertex, the number of these paths that pass through a specific node. This property in particular is important in the context of biological pathways, since it can detect the existence of hubs and highly connected nodes even within a large network. Mathematically, it is defined as:

$$g_L = \sum_{L \neq i \neq j} \frac{\sigma_{ij}(L)}{\sigma_{ij}} \tag{3}$$

where L is the node under analysis, σ_{ij} is the number of shortest paths between i and j and $\sigma_{ij}(L)$ is the number of shortest paths passing through L

The final analyzed subnet had a total of 4868 nodes and 74123 edges. The alpha and beta subunits showed respectively a connectivity of 57 and 42 nodes; both had a transitivity of 0.313. Lastly, the betweenness centrality had a value of 0.0107 for the alpha subunit and of 0.004 for the beta subunit.

These results alone, however, are not very indicative of the kind of proteins hemoglobin interacts with and in which pathways it might implicated. In order to further examine the outcome of this analysis reducing the amount of computational effort, the network including only proteins that had a direct interaction with the two subunits (path length = 1) was computed. For each of these proteins, the degree, transitivity and betweenness centrality in the network with maximum path length 2 was calculated.
Only the top ten results, together with the values of the two hemoglobin subunits, are reported here (Table 7).

From this analysis, it emerged that the two hemoglobin subunits are not the most intertwined proteins of the network: vascular cell adhesion protein 1 shows the highest degree value (641), followed by proliferation marker protein Ki-67, heat shock cognate 71kDa protein and mitogen-activated protein kinase 6, all with degrees above 400.

The degree of connectivity is consistent with the biological role these proteins play in cell processes: vascular cell adhesion protein 1 is important in cell-cell recognition but also mediates signal transduction. The second and the fourth are implicated in the cell cycle, while the third is a molecular chaperone.
All these functions require a high level of regulation and the fine tuning of expression of the proteins involved. For this reason, these proteins must interact with a great number of other polypeptides in order not to disrupt essential cell functions.

Proteins showing high transitivity in the network are other vari-

Figure 8 The protein subnetwork extracted form the IntAct database comprising only proteins with a maximum path length of two from the α and β hemoglobin subunits (indicated in red).

ants of the hemoglobin complex, i.e. epsilon (HBE), delta (HBD), theta-1 (BHQ1) and gamma-2 (HBG2). The first is a beta-type chain, expressed early in mammalian embryos; HBD is the most common variant for the subunit beta in adults and, together with the HBA forms HbA-2, which represents the 3% of adult hemoglobin. BHQ1 mRNA is found in human fetal erythroid tissue, while HBG2 is expressed in fetal liver, spleen and bone marrow and normally is subsequently replaced by HbA.
The high clustering coefficient of these variants might be due to the way the subnet was constructed, since the whole proteome was firstly filtered for proteins related in some way with the standard hemoglobin subunits in order to reduce the amount of computational calculations.

Finally, the top three and the fifth proteins per betweenness centrality are the ones already found when considering the highest degree. This might be explained by the fact that a node which is highly connected to the others in the graph, it also often crossed when searching for the shortest possible path to connect two distant vertexes. Hence, in some way, these two properties might be interconnected.

DISCUSSION

In this study, proof of intra and interchain interactions within and between the four monomers that make up the hemoglobin complex have been obtained through a bioinformatical approach.
An high symmetry existing within the complex was highlighted by all the analysis carried out in this study.
The detection of histidine residues anchoring the hetero group to the single chains through interactions both with the iron and the oxygen atoms sheds light on the importance these amino acids play in the stabilization of the association. Moreover, the high percentage of apolar residues close by the heme group confirms the theory that the porphyrin ring inserts itself in a hydrophobic pocket found within the folded protein. In particular, the heme group inserts its apolar upper part into the polypeptide, leaving the carboxyl group outside in direct contact with the solvent.
For what concerns interchain interactions, insights on the non-covalent bonds existing between the monomers have been obtained. In particular, hydrophobic hotspots seem to contribute for the major part to holding the $\alpha\beta$ dimer together, while stabilization of the tetrameric complex occurs mainly thanks to the existance of numerous salt bridges between chain A and C. These results are also in concordance with previous findings that emphasized the major role non-covalent bonds play in the maintenance of the T state of deoxyhemoglobin.
In conclusion, an in-depth analysis of the interconnectivity of the alpha and beta subunit in the human proteome network was performed. The main results highlight how these two proteins do not form hubs or have high degree values if compared with other members of the subnet.

LITERATURE CITED

Bettati, S., A. Mozzarelli, and M. F. Perutz, 1998 Allosteric mechanism of haemoglobin: Rupture of salt-bridges raises the oxygen affinity of the t-structures. Journal of Molecular Biology .

Table 7 The degree, clustering coefficient and betweenness centrality for the top ten results emerged from this analysis.
To each UniProt ID, the protein name has been associated. Full tables for all the 82 proteins in the network on Supplementary materials.

Position	UniProt ID	Degree	Protein name
1	P19320	641	Vascular cell adhesion protein 1
2	P46013	443	Proliferation marker protein Ki-67
3	P11142	425	Heat shock cognate 71 kDa protein
4	Q16659	405	Mitogen-activated protein kinase 6
5	Q15323	398	Keratin, type I cuticular Ha1
6	Q6A162	386	Keratin, type I cytoskeletal 40
7	Q7Z3S9	369	Notch homolog 2 N-terminal-like protein A
8	P27348	344	14-3-3 protein theta
9	P36957	279	Component of 2-oxoglutarate dehydrogenase complex, mitochondrial
10	P30480	271	HLA class I histocompatibility antigen, B alpha chain
32	P69905	57	Hemoglobin subunit alpha
41	P68871	42	Hemoglobin subunit beta
Position	UniProt ID	Transitivity	Protein name
1	P02100	0.4	Hemoglobin subunit epsilon
2	P09105	0.25	Hemoglobin subunit theta-1
3	P02042	0.2143	Hemoglobin subunit delta
4	Q9BQ69	0.166	ADP-ribose glycohydrolase MACROD1
5	Q15904	0.1558	V-type proton ATPase subunit S1
6	P69892	0.1455	Hemoglobin subunit gamma-2
7	Q9HCI7	0.1364	E3 ubiquitin-protein ligase MSL2
8	Q9HBW0	0.1333	Lysophosphatidic acid receptor 2
9	Q9UNS2	0.1301	COP9 signalosome complex subunit 3
10	O75293	0.1238	Growth arrest and DNA damage-inducible protein GADD45 beta
38	P68871	0.0314	Hemoglobin subunit beta
39	P69905	0.0313	Hemoglobin subunit alpha
Position	UniProt ID	Betweenness centrality	Protein name
1	P11142	0.046	Heat shock cognate 71 kDa protein
2	P19320	0.0312	Vascular cell adhesion protein 1
3	Q16659	0.0299	Mitogen-activated protein kinase 6
4	Q15323	0.0238	Keratin, type I cuticular Ha1
5	P46013	0.0217	Proliferation marker protein Ki-67
6	Q7Z3S9	0.0211	Notch homolog 2 N-terminal-like protein A
7	Q6A162	0.0207	Keratin, type I cytoskeletal 40
8	P05067	0.0171	Amyloid-beta precursor protein
9	P27348	0.015	14-3-3 protein theta
10	Q15051	0.0122	IQ calmodulin-binding motif-containing protein 1
12	P69905	0.0107	Hemoglobin subunit alpha
25	P68871	0.0038	Hemoglobin subunit beta

Bosshard, H. R., D. N. Marti, and I. Jelesarov, 2004 Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. Journal of Molecular Recognition .

Felicetti, L., B. Colombo, and C. Baglioni, 1966 Assembly of hemoglobin. Biochimica et Biophysica Acta .

Jensen, F. B., A. Fago, and R. E. Weber, 1998 Hemoglobin structure and function .

Keskin, O., A. Gursoy, B. Ma, and R. Nussinov, 2007 Principles of protein-protein interactions: What are the preferred ways for proteins to interact? .

Morrow, J. K. and S. Zhang, 2013 Computational prediction of hot spot residues .

Mrabet, N. T., M. J. McDonald, S. Turci, R. Sarkar, A. Szabos, *et al.*, 1985 Electrostatic attraction governs the dimer assembly of human hemoglobin. The Journal of Biological Chemistry .

Paoli, M., R. Liddington, J. Tame, A. Wilkinson, and G. Dodson, 1996 Crystal structure of t state haemoglobin with oxygen bound at all four haems. Journal of Molecular Biology .

Perutz, M. F., 1990 Mechanisms regulating the reactions of human hemoglobin with oxygen and carbon monoxide .

Supplementary materials

Hemoglobin complex analysis: structural properties and interactions within the human proteome

Del Missier Giorgia

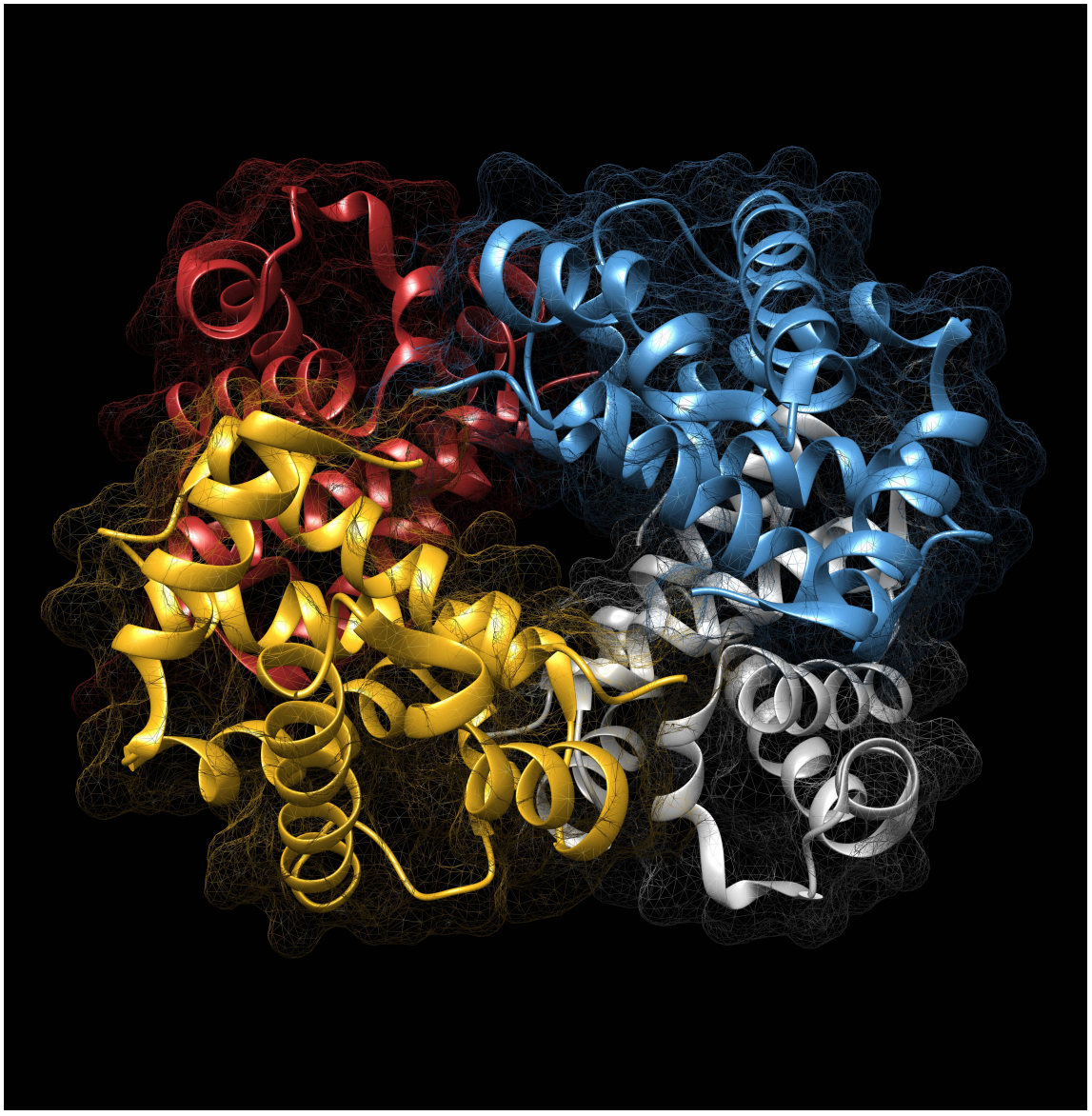

Figure 1: The whole hemoglobin quaternary complex, as visualized by the Chimera software.
Chain A (blue), Chain B (gray), Chain C (gold), Chain D (red).

Complete tables, for each chain, of residues found at a maximum distance of 3.5Å from the heme group.

Chain	Residue	Hetero	Best distance	Atoms ($\leq 3.5\AA$)
A	HIS58	HEM1142	3.331724	CE1-CHA, CE1-C1A, CE1-C4D
A	PHE98	HEM1142	3.330067	CE1-CHC, CD1-CBB
A	PHE43	HEM1142	3.381986	CZ-CHD, CE2-CMD
A	LEU83	HEM1142	3.4444	CD1-C3A
A	HIS87	HEM1142	2.259166	CE1-C4A, NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
A	LYS61	HEM1142	3.277009	O-CMA, CE-O1A
A	LEU86	HEM1142	3.365984	CD2-CBA
A	ASN97	HEM1142	3.169213	O-CMC
A	VAL93	HEM1142	3.433899	CG1-CAC
A	LEU91	HEM1142	3.496219	CD1-C4D
A	TYR42	HEM1142	3.341404	O-CMD
A	PHE46	HEM1142	3.409717	CE2-O1D
A	HIS45	HEM1142	3.066142	NE2-O2D
A	HIS58	OXY1143	2.835774	NE2-O1, NE2-O2
A	VAL62	OXY1143	3.3549	CG2-O2

Table 1: Heme - Monomer interactions, Chain A

Chain	Residue	Hetero	Best distance	Atoms ($\leq 3.5\AA$)
B	HIS206	HEM1290	3.238752	CE1-CHA, CE1-C1A, NE2-C4D
B	LEU234	HEM1290	3.454745	CD2-CBA
B	LEU284	HEM1290	3.43028	CD1-CAB
B	ASN245	HEM1290	3.327893	O-CMC
B	LEU239	HEM1290	3.410092	CD1-C3D, CD1-C4D
B	HIS235	HEM1290	2.161144	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
B	HIS206	OXY1291	2.531888	NE2-O1, NE2-O2
B	VAL210	OXY1291	2.865797	CG2-O1, CG2-O2

Table 2: Heme - Monomer interactions, Chain B

Chain	Residue	Hetero	Best distance	Atoms ($\leq 3.5\AA$)
C	PHE498	HEM1542	3.297848	CE1-CHC
C	LYS461	HEM1542	3.202792	O-CMA, NZ-O1D
C	ASN497	HEM1542	3.347145	O-CMC, CB-CBC
C	VAL493	HEM1542	3.439595	CG1-CAC
C	LEU491	HEM1542	3.48939	CD1-C4D
C	TYR442	HEM1542	3.296493	O-CMD
C	HIS445	HEM1542	2.947272	NE2-O2D
C	HIS487	HEM1542	2.352667	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
C	HIS458	OXY1543	2.471012	NE2-O1, NE2-O2
C	VAL462	OXY1543	3.344702	CG2-O2,

Table 3: Heme - Monomer interactions, Chain C

Chain	Residue	Hetero	Best distance	Atoms ($\leq 3.5\mathring{A}$)
D	PHE646	HEM1690	3.470102	CD1-CHC
D	LEU649	HEM1690	3.089558	CD2-CHC, CD2-CBB, CD2-CMC
D	LYS609	HEM1690	2.667527	NZ-CAA, NZ-O1A
D	LEU684	HEM1690	3.145259	CD1-C3B, CD1-CAB
D	ASN645	HEM1690	3.397383	O-CMC
D	LEU639	HEM1690	3.221423	CD1-C3D, CD1-C4D
D	HIS635	HEM1690	2.223512	NE2-NA, NE2-NB, NE2-NC, NE2-ND, NE2-FE
D	HIS606	OXY1691	2.757016	NE2-O1, NE2-O2
D	VAL610	OXY1691	2.440367	CG2-O1, CG2-O2

Table 4: Heme - Monomer interactions, Chain D

Complete tables of residues found at a maximum distance of 3.5Å for each combination of the four monomers in the hemoglobin complex.

Residue	Chain 2	Residue 2	Best distance	Atoms ($\leq 3.5\AA$)
ARG31	B	GLN270	3.190034	CB-OE1, CD-OE1, NH1-OE1
ARG31	B	PHE265	2.998073	CZ-O, NH1-O, NH2-O
ARG31	B	THR266	3.355993	NH1-O
LEU34	B	ALA271	3.35179	O-CB
SER35	B	GLN274	3.397808	CB-CG
SER35	B	GLN270	3.404292	OG-CD
HIS103	B	ASN251	3.347603	CE1-OD1
HIS103	B	GLN274	2.854999	NE2-OE1
ALA111	B	GLY262	3.342599	O-CA
HIS112	B	LYS263	3.341458	CE1-NZ
PRO114	B	HIS259	2.200743	CA-NE2, C-NE2, O-NE2
PHE117	B	ARG173	3.270548	O-NH2, CD2-NH2, CE2-NH2
PHE117	B	HIS259	3.37739	O-NE2
HIS122	B	ARG173	3.047758	CB-NH2, ND1-NH2
ASP126	C	ARG541	3.240818	CG-NH1, OD2-NH2
LYS127	C	ARG541	3.476898	NZ-OXT
ARG141	C	LYS527	2.386822	C-NZ, O-NZ, OXT-NZ
ARG141	C	ASP526	2.814262	CZ-OD2, NH1-OD2, NH2-OD2
LYS40	D	HIS689	3.21761	CE-OXT, NZ-OXT
THR41	D	HIS640	3.403172	O-CB
TYR42	D	ARG583	3.451242	CD1-NH1, CE1-NH1
TYR42	D	ASP642	2.756334	OH-OD1
LEU91	D	ARG583	2.968311	O-NH2
ARG92	D	ARG583	3.398686	O-CD, CB-CD, CZ-CB, NH1-CA
ARG92	D	GLU586	2.653403	NH1-OE1
ASP94	D	TRP580	3.497119	OD1-NE1
ASN97	D	ASP642	2.894801	OD1-OD1
ARG141	D	VAL577	2.898728	NE-O, CZ-O, NH1-CG1
ARG141	D	TYR578	3.492478	NH1-CE1

Table 5: Interactions between monomers, Chain A

Residue	Chain 2	Residue 2	Best distance	Atoms ($\leq 3.5\mathring{A}$)
ARG173	A	PHE117	2.734847	NH1-O, NH2-CE2
ARG173	A	HIS122	3.047758	NH2-ND1
ASN251	A	HIS103	3.347603	OD1-CE1
HIS259	A	PRO114	2.200743	CD2-O, CE1-O, NE2-O
HIS259	A	PHE117	3.37739	NE2-O
GLY262	A	ALA111	3.342599	CA-O
LYS263	A	HIS112	3.341458	NZ-CE1
PHE265	A	ARG31	2.998073	O-NH2
THR266	A	ARG31	3.355993	O-NH1
GLN270	A	SER35	3.404292	CD-OG
GLN270	A	ARG31	3.190034	OE1-NH1
ALA271	A	LEU34	3.35179	CB-O
GLN274	A	SER35	3.397808	CG-CB
GLN274	A	HIS103	2.854999	OE1-NE2
VAL177	C	ARG541	3.073421	O-NH1, CG1-NH1
TRP180	C	ASP494	3.32044	NE1-OD1
TRP180	C	ARG541	3.375408	CZ3-CG
ARG183	C	ARG492	3.40121	CD-O
ARG183	C	LEU491	3.357893	NH2-O
HIS240	C	THR441	3.333462	O-CB, CB-O
ASP242	C	TYR442	2.23775	CB-OH, CG-OH, OD1-OH
ASP242	C	ASN497	2.993975	OD1-OD1
PRO243	C	THR438	3.318125	CG-CG2
GLU244	C	ASP494	3.258756	CG-OD2
HIS289	C	LYS440	2.694048	OXT-NZ

Table 6: Interactions between monomers, Chain B

Residue	Chain 2	Residue 2	Best distance	Atoms ($\leq 3.5\mathring{A}$)
ASP526	A	ARG141	3.244306	OD2-NH2
LYS527	A	ARG141	3.025898	CD-O, CE-O, NZ-OXT
ARG541	A	LYS127	2.775127	C-NZ, O-NZ, OXT-NZ
ARG541	A	ASP126	2.44018	CZ-OD2, NH1-OD2, NH2-OD2,
THR438	B	PRO243	3.318125	CG2-CG
LYS440	B	HIS289	2.694048	NZ-OXT
THR441	B	HIS240	3.333462	O-CB, CB-O
TYR442	B	ASP242	2.23775	CE2-OD1, CZ-OD1, OH-OD1
LEU491	B	ARG183	3.357893	O-NH2
ARG492	B	ARG183	3.40121	O-CD
ASP494	B	TRP180	3.32044	OD1-NE1
ASP494	B	GLU244	3.258756	OD1-CG, OD2-CG
ASN497	B	ASP242	2.993975	OD1-OD1
ARG541	B	TRP180	3.375408	CG-CZ3
ARG541	B	VAL177	3.127981	NE-O, NH1-CG1
ARG431	D	GLN670	2.554207	CB-OE1, CG-OE1, CD-OE1, CZ-OE1, NH1-OE1
ARG431	D	PHE665	2.762317	CZ-O, NH1-O, NH2-O
SER435	D	GLN670	3.44066	OG-NE2
HIS503	D	GLN674	2.94307	NE2-OE1
VAL507	D	GLN670	3.374879	CG1-OE1
ALA510	D	HIS659	3.46821	O-CD2
ALA511	D	ALA658	3.476256	CA-O
ALA511	D	GLY662	3.287494	O-CA
PRO514	D	HIS659	2.489002	CA-NE2, C-NE2, O-NE2
PHE517	D	ARG573	3.111129	O-NH2, CD2-NH2, CE2-NH2
PHE517	D	HIS659	3.405998	O-NE2
HIS522	D	ARG573	2.836934	CB-NH2, CG-NH2, ND1-NH2

Table 7: Interactions between monomers, Chain C

Residue	Chain 2	Residue 2	Best distance	Atoms ($\leq 3.5\mathring{A}$)
VAL577	A	ARG141	2.581993	O-NH1, CG1-NH1
TYR578	A	ARG141	3.492478	CE1-NH1
TRP580	A	ASP94	3.497119	NE1-OD1
ARG583	A	ARG92	3.398686	CA-NH1, CB-CZ, CD-CB
ARG583	A	TYR42	3.463058	NH1-CE1
ARG583	A	LEU91	2.968311	NH2-O
GLU586	A	ARG92	2.653403	OE1-NH1
HIS640	A	THR41	3.403172	CB-O
ASP642	A	TYR42	2.756334	CB-OH, CG-OH, OD1-OH
ASP642	A	ASN97	2.894801	OD1-OD1
HIS689	A	LYS40	3.21761	OXT-NZ
ARG573	C	HIS522	2.836934	CD-ND1, NH2-ND1
ARG573	C	PHE517	2.781586	CZ-O, NH1-O, NH2-CE2
ALA658	C	ALA511	3.476256	O-CA
HIS659	C	ALA510	3.46821	CD2-O
HIS659	C	PRO514	2.489002	CE1-O, NE2-O
HIS659	C	PHE517	3.29276	CE1-O, NE2-O
GLY662	C	ALA511	3.287494	CA-O
PHE665	C	ARG431	2.762317	O-NH2
GLN670	C	ARG431	2.554207	CB-NH1, OE1-NH1
GLN670	C	VAL507	3.374879	OE1-CG1
GLN670	C	SER435	3.44066	NE2-OG
GLN674	C	HIS503	2.94307	OE1-NE2

Table 8: Interactions between monomers, Chain D

Complete tables of residues that lost more than 10% of relative solvent accessible area in the complex if compared to the RASA they show when in the single chain.

Chain	Residue	RSA(M)	RSA(C)	RSA(M) - RSA(C)
A	GLU30	0.3211	0.2053	0.1158
A	ARG31	0.4089	0.0178	0.3911
A	LEU34	0.7588	0.4471	0.3118
A	SER35	0.687	0.1391	0.5478
A	PRO37	0.5448	0.3172	0.2276
A	THR38	0.6714	0.2714	0.4
A	LYS40	0.44	0.32	0.12
A	THR41	0.7571	0.0571	0.7
A	TYR42	0.3174	0.113	0.2043
A	PRO44	0.731	0.469	0.2621
A	ARG92	0.7467	0.1911	0.5556
A	ASP94	0.58	0.0533	0.5267
A	VAL96	0.5806	0.3161	0.2645
A	HIS103	0.5179	0.0872	0.4308
A	VAL107	0.3226	0.0	0.3226
A	ALA110	0.3043	0.0	0.3043
A	ALA111	0.5565	0.087	0.4696
A	HIS112	0.3179	0.1641	0.1538
A	PRO114	0.6828	0.4483	0.2345
A	PHE117	0.1905	0.0095	0.181
A	PRO119	0.7655	0.1448	0.6207
A	ALA120	0.4609	0.2783	0.1826
A	HIS122	0.4308	0.0103	0.4205
A	ALA123	0.4957	0.1478	0.3478
A	ASP126	0.5333	0.14	0.3933
A	LYS127	0.425	0.125	0.3
A	SER138	0.6348	0.513	0.1217
A	TYR140	0.2739	0.0826	0.1913
A	ARG141	1	0.3422	0.6578

Table 9: Loss of relative solvent accessibility, Chain A

Chain	Residue	RSA(M)	RSA(C)	RSA(M) - RSA(C)
B	ARG173	0.4089	0.0222	0.3867
B	VAL176	0.3161	0.1226	0.1935
B	VAL177	0.5548	0.0065	0.5484
B	TYR178	0.2826	0.1261	0.1565
B	PRO179	0.4483	0.2414	0.2069
B	TRP180	0.6235	0.0745	0.549
B	ARG183	0.7333	0.2933	0.44
B	PRO194	0.5103	0.3241	0.1862
B	HIS240	0.6923	0.3385	0.3538
B	ASP242	0.6267	0.0133	0.6133
B	PRO243	0.1724	0.0	0.1724
B	GLU244	0.4368	0.1684	0.2684
B	ASN251	0.4875	0.3	0.1875
B	CYS255	0.6148	0.0222	0.5926
B	ALA258	0.3304	0.0	0.3304
B	HIS259	0.6513	0.1846	0.4667
B	GLY262	0.6133	0.24	0.3733
B	LYS263	1	0.725	0.275
B	PRO267	0.6966	0.1103	0.5862
B	PRO268	0.6138	0.4828	0.131
B	GLN270	0.45	0.0	0.45
B	ALA271	0.4435	0.087	0.3565
B	GLN274	0.3389	0.0833	0.2556
B	HIS286	0.4923	0.3744	0.1179
B	TYR288	0.2522	0.1261	0.1261
B	HIS289	0.9692	0.6205	0.3487

Table 10: Loss of relative solvent accessibility, Chain B

Chain	Residue	RSA(M)	RSA(C)	RSA(M) - RSA(C)
C	VAL401	1	0.8452	0.1548
C	GLU430	0.3474	0.2368	0.1105
C	ARG431	0.4267	0.0311	0.3956
C	LEU434	0.7529	0.4765	0.2765
C	SER435	0.6522	0.1304	0.5217
C	PHE436	0.3	0.1952	0.1048
C	PRO437	0.531	0.331	0.2
C	THR438	0.6786	0.2857	0.3929
C	LYS440	0.445	0.29	0.155
C	THR441	0.7286	0.0643	0.6643
C	TYR442	0.3304	0.113	0.2174
C	PRO444	0.7655	0.5103	0.2552
C	ARG492	0.7333	0.2044	0.5289
C	ASP494	0.5333	0.0467	0.4867
C	PRO495	0.4345	0.3241	0.1103
C	VAL496	0.5806	0.3226	0.2581
C	HIS503	0.5026	0.0872	0.4154
C	VAL507	0.3355	0.0	0.3355
C	ALA510	0.2957	0.0	0.2957
C	ALA511	0.5565	0.087	0.4696
C	PRO514	0.6759	0.4621	0.2138
C	PHE517	0.2048	0.0095	0.1952
C	PRO519	0.7241	0.1103	0.6138
C	ALA520	0.5391	0.287	0.2522
C	HIS522	0.4462	0.0103	0.4359
C	ALA523	0.513	0.1217	0.3913
C	ASP526	0.5333	0.1533	0.38
C	LYS527	0.405	0.095	0.31
C	TYR540	0.2522	0.0783	0.1739
C	ARG541	1	0.3556	0.6444

Table 11: Loss of relative solvent accessibility, Chain C

Chain	Residue	RSA(M)	RSA(C)	RSA(M) - RSA(C)
D	HIS545	0.9179	0.7949	0.1231
D	ARG573	0.4044	0.0133	0.3911
D	VAL576	0.3484	0.1226	0.2258
D	VAL577	0.529	0.0	0.529
D	TYR578	0.2696	0.113	0.1565
D	PRO579	0.4483	0.2552	0.1931
D	TRP580	0.6196	0.0745	0.5451
D	ARG583	0.7156	0.2622	0.4533
D	GLU586	0.7737	0.6263	0.1474
D	PRO594	0.4759	0.2552	0.2207
D	HIS640	0.6923	0.3641	0.3282
D	ASP642	0.64	0.0133	0.6267
D	PRO643	0.1517	0.0276	0.1241
D	GLU644	0.4789	0.2158	0.2632
D	ASN651	0.4688	0.2938	0.175
D	CYS655	0.6519	0.0222	0.6296
D	ALA658	0.3304	0.0	0.3304
D	HIS659	0.6205	0.1692	0.4513
D	GLY662	0.56	0.16	0.4
D	PRO667	0.6621	0.1103	0.5517
D	PRO668	0.6069	0.4897	0.1172
D	GLN670	0.4444	0.0	0.4444
D	ALA671	0.4261	0.0957	0.3304
D	GLN674	0.35	0.0667	0.2833
D	TYR688	0.2652	0.1087	0.1565
D	HIS689	0.9641	0.6462	0.3179

Table 12: Loss of relative solvent accessibility, Chain D

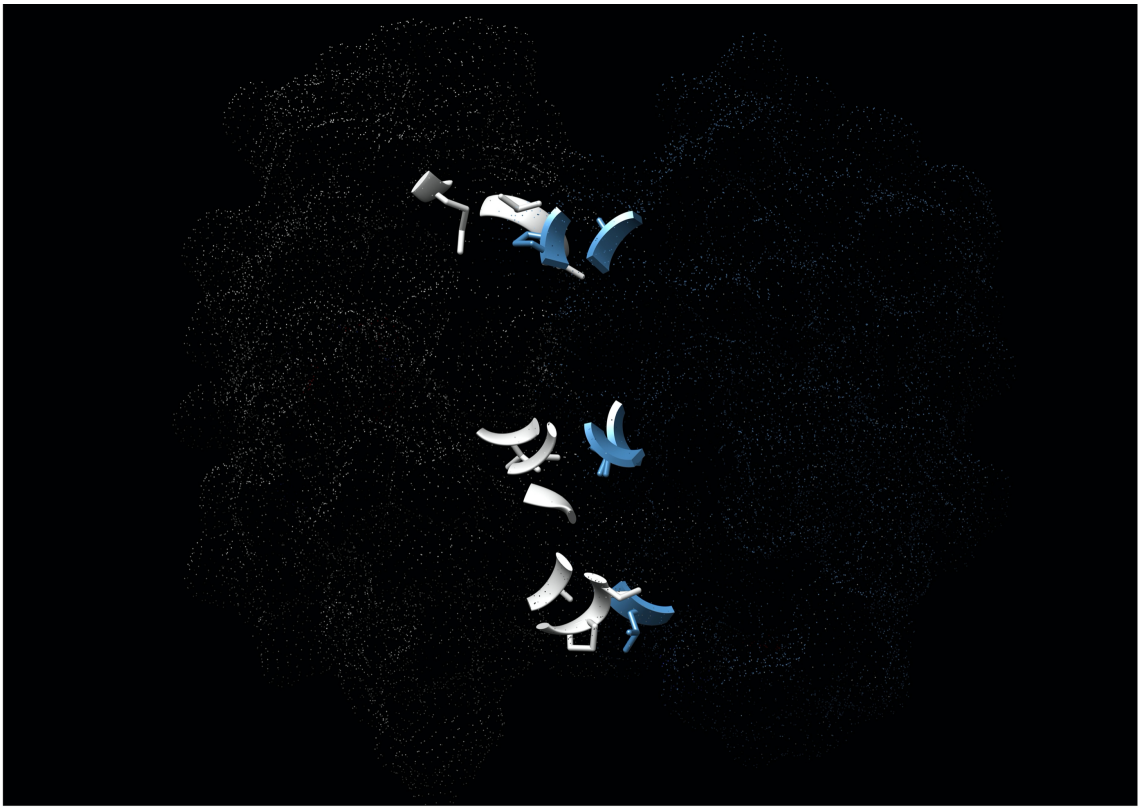

Figure 2: Chain A (blue) and chain B (gray). Residues predicted to form hotspots, important in the stabilization of the $\alpha\beta$ dimer.

Complete tables illustrating the degree, transitivity and betweenness centrality of the proteins having direct connections with the α and β subunit of hemoglobin. The network comprises 80 protein, plus the two subunits under analysis.

Position	UniProt ID	Degree
1	P19320	641
2	P46013	443
3	P11142	425
4	Q16659	405
5	Q15323	398
6	Q6A162	386
7	Q7Z3S9	369
8	P27348	344
9	P36957	279
10	P30480	271
11	O76011	264
12	P56945	256
13	P05067-PRO$_0$000000092	220
14	P15336	215
15	Q15051	212
16	Q99714	185
17	O60506	162
18	P62195	124
19	Q9UER7	118
20	P55209	106
21	Q9UHK0	91
22	Q9H8J5	89
23	P02647	76
24	P00387	73
25	P0DMV8	65
26	Q96T60	64
27	P20036	64
28	Q9UNS2	63
29	P55899	61
30	Q9H3Z4	61
31	Q09019	58
32	P69905	57
33	Q8N878	47
34	O15151	47
35	Q05086	46
36	P60484	46
37	Q92574	46
38	Q13268	45
39	Q2TAC2	44
40	Q5TEU4	43
41	P68871	42

Position	UniProt ID	Degree
42	A2RU67	41
43	P32971	38
44	P29474	36
45	P01871	35
46	P12314	33
47	P32456	31
48	P02008	30
49	P32189-1	27
50	Q8N9F7-2	27
51	Q02383	25
52	Q15904	24
53	Q9BQ69	23
54	O75293	21
55	Q9UI36-2	21
56	Q9H1J1-2	20
57	Q8TAC2	19
58	P07339	19
59	P06899	18
60	Q969P0-3	16
61	O95156	15
62	P30711	13
63	Q6UXU6	12
64	Q9HCI7	12
65	P62341	11
66	P69892	11
67	Q9NZD4	10
68	Q8J025	10
69	P09105	9
70	O43157-2	9
71	O60759	9
72	Q8N8W4-2	8
73	P02042	8
74	Q9NXJ5	7
75	Q8TDB8-2	6
76	P20933	6
77	P02100	6
78	Q9HBW0	6
79	Q9UQ53-2	4
80	P22003	4
81	Q6UWM9	4
82	Q9ULK0	4

Table 13: Degree, in descending order

Position	UniProt ID	Transitivity
1	P02100	0.4
2	P09105	0.25
3	P02042	0.2143
4	Q9BQ69	0.166
5	Q15904	0.1558
6	P69892	0.1455
7	Q9HCI7	0.1364
8	Q9HBW0	0.1333
9	Q9UNS2	0.1301
10	O75293	0.1238
11	Q9H1J1-2	0.1105
12	Q9NXJ5	0.0952
13	P62195	0.0902
14	Q5TEU4	0.0819
15	Q8N9F7-2	0.0741
16	Q8N8W4-2	0.0714
17	O15151	0.0694
18	P20933	0.0667
19	Q8TDB8-2	0.0667
20	Q99714	0.0662
21	P02008	0.0621
22	P29474	0.0619
23	O43157-2	0.0556
24	P55209	0.051
25	Q92574	0.0465
26	P06899	0.0458
27	P02647	0.0415
28	P27348	0.0409
29	P01871	0.0398
30	Q9UER7	0.0393
31	P00387	0.0392
32	Q13268	0.0364
33	P07339	0.0351
34	Q8TAC2	0.0351
35	P36957	0.035
36	Q7Z3S9	0.0346
37	P0DMV8	0.0327
38	P68871	0.0314
39	P69905	0.0313
40	P60484	0.0296
41	Q05086	0.0285
42	P15336	0.0275
43	O60506	0.0271
44	P12314	0.0265
45	Q969P0-3	0.025
46	Q09019	0.0248
47	Q96T60	0.0248
48	O76011	0.0228
49	Q15051	0.0227
50	P46013	0.0201
51	P11142	0.0193
52	P62341	0.0182
53	Q9UI36-2	0.0175
54	P55899	0.0175
55	Q2TAC2	0.0163
56	P19320	0.0161
57	Q9H3Z4	0.0158
58	Q15323	0.0155
59	P56945	0.0141
60	P32456	0.0129
61	P30711	0.0128
62	P30480	0.0126
63	Q16659	0.0122
64	P32189-1	0.0114
65	Q6A162	0.0109
66	Q02383	0.01
67	P05067-PRO$_0$000000092	0.0089
68	A2RU67	0.0085
69	P20036	0.0084
70	Q9UHK0	0.0064
71	P32971	0.0057
72	Q9H8J5	0.0056
73	Q8N878	0.0037
74	Q9ULK0	0
75	Q8J025	0
76	Q6UXU6	0
77	P22003	0
78	Q6UWM9	0
79	Q9NZD4	0
80	Q9UQ53-2	0
81	O60759	0
82	O95156	0

Table 14: Transitivity, in descending order

Position	UniProt ID	BC
1	P11142	0.046
2	P19320	0.0312
3	Q16659	0.0299
4	Q15323	0.0238
5	P46013	0.0217
6	Q7Z3S9	0.0211
7	Q6A162	0.0207
8	P05067-PRO$_0$000000092	0.0171
9	P27348	0.015
10	Q15051	0.0122
11	O76011	0.011
12	P69905	0.0107
13	O60506	0.0089
14	P56945	0.0084
15	Q9UHK0	0.0081
16	P15336	0.0074
17	P62195	0.0063
18	Q9H8J5	0.0061
19	P36957	0.006
20	P30480	0.0059
21	P02647	0.0046
22	Q9UER7	0.0043
23	P0DMV8	0.0043
24	P00387	0.0039
25	P68871	0.0038
26	P20036	0.0034
27	Q9H3Z4	0.0033
28	P55899	0.0032
29	Q99714	0.0028
30	Q13268	0.0027
31	Q96T60	0.0026
32	O15151	0.0023
33	P55209	0.0022
34	Q8N878	0.0021
35	P32971	0.0019
36	A2RU67	0.0016
37	P12314	0.0015
38	Q9UNS2	0.0015
39	Q92574	0.0013
40	P01871	0.0012
41	Q09019	0.0012

Position	UniProt ID	BC
42	Q5TEU4	0.0011
43	Q8TAC2	0.001
44	Q05086	0.0009
45	Q02383	0.0009
46	P02008	0.0008
47	Q6UWM9	0.0008
48	P60484	0.0007
49	Q969P0-3	0.0006
50	P32456	0.0005
51	P32189-1	0.0005
52	O43157-2	0.0004
53	P09105	0.0004
54	Q2TAC2	0.0004
55	P29474	0.0004
56	Q9HCI7	0.0003
57	Q8N9F7-2	0.0003
58	Q9H1J1-2	0.0003
59	O75293	0.0002
60	O95156	0.0002
61	Q9UI36-2	0.0002
62	Q15904	0.0002
63	Q9NXJ5	0.0002
64	P06899	0.0002
65	Q6UXU6	0.0002
66	P07339	0.0001
67	P69892	0.0001
68	P30711	0.0001
69	P02042	0.0001
70	Q9NZD4	0.0001
71	Q9BQ69	0.0001
72	P62341	0.0001
73	P22003	0.0001
74	Q8J025	0.0001
75	O60759	0.0001
76	Q8TDB8-2	0.0
77	Q8N8W4-2	0.0
78	P20933	0.0
79	P02100	0.0
80	Q9UQ53-2	0.0
81	Q9HBW0	0.0
82	Q9ULK0	0.0

Table 15: Betweenness centrality, in descending order