By Amelia Kurti

Structural Analysis of the Hemoglobin HbA Complex

The essential role of Haemoglobin as an Oxygen transporter has pushed scientists to thoroughly study its structure for decades. However, the crucial protein seems to not have fully uncovered its functional potential yet (Hem research). This report aims to present the analysis of the Haemoglobin HbA complex structure, performed by utilising combined bioinformatical tools to understand which residues are responsible for the formation of the protein complex, which residues hold together the tetramer and provide it with the Oxygen transporter function, as well as the protein-protein interaction involvement of α and β subunits and their position in Human Interactome. Files retrieved from major databases as PDB and IntAct are adjusted, parsed and analysed through LINUX and Python. UCSF Chimera is used for verification of acquired results regarding residues interactions.
Haemoglobin is an astonishing molecule, with subunits of the same composition, yet diverse features. The evidence proposes that chains conserve the composition of the haem pocket, with the proximal and distal Histidine always present, due to their role in ligand binding. A slightly different mechanism of ligand stabilization is displayed by chains of subunit B, where chain B appears to have no residue stabilizing the protoporphyrin tails, instead, this is accomplished by intracellular ions. Rather than stabilising the propionate tails by a Histidine like chains A and C, chain D has chosen Lysine, another positively charged amino acid. Another interesting difference, in oxy T state the interaction between chains C and D, computationally, conveys the impression of being stronger and involves more area of the monomers compared to chains A and B interaction. Subunits α and β show as well as different connectivity and interactivity in the human interactome, with α interacting with regulatory proteins and β, able to network only with other haemoglobin subunits like subunit zeta and epsilon. What is the causative factor of these variations within and between subunits composition and the role in protein interaction network is still on the study, since the complexity of Haemoglobin suggests to be broader than the scientific community expected.

INTRODUCTION:

The Globin protein Superfamily, extended in all life domains, is marked by 6-8 α helixes conferring its globular structure and functional properties, and the presence of a haem prosthetic group, which promotes the binding and transport of gaseous ligands such as O_2, CO, NO[1]. The intensive study of this superfamily commences with the resolution of Myoglobin structure, extracted from Sperm Whale muscles, using X-ray Crystallography in 1958 (A 3D). Accordingly, Myoglobin is used as a structural and functional reference for the rest of the Globin superfamily members. Further in-depth studies indicate that an ancestral Globin evolved to give rise to Neuroglobin, Cytoglobin, Myoglobin and Haemoglobin hundreds of million years ago.[2](**Figure 1a**)
From Bacteria to Vertebrates, the Globins display different functionalities [2]. In bacteria, archaea and fungi, the Globins are identified as Flavohemoglobins, having primary importance the NO detoxification[3]; in distinct taxa, including protists and plant, globin proteins of single-chain store and transport O_2, and transfer electrons, whereas in Vertebra, the multi-chain globin proteins evolved to grant long-distance transport of diatomic gases[2]. Globin sequence alignment studies have concluded that the 3D structure of Globins is well preserved while the sequences are quite distinct[1]. The evolution has spared from altering only a single residue HisF8, found 100% conserved over the whole superfamily.
An iron proto-porphyrin-XI (haem) is bound to every Globin peptide. The amphipathic nature of the haem molecule is delivered by two charged propionate groups which interact with water and/or polar amino acids side chains of the globin surface, and the remaining highly hydrophobic part of the molecule, interacting with the internal hydrophobic residues of globin[4]. Each of the four pyrrole rings of porphyrin contributes with one N ligand in the stabilisation of the central iron of the haem. Iron is kept in a ferrous state, necessary to reversibly bind to molecular oxygen. The binding of the haem to the globin peptide is attributed to noncovalent interactions between porphyrin and globin, and by a covalent bond of the central iron with the imidazole chain of HisF8. This region of the haem pocket is known as the proximal haem pocket, while on the opposite side of the porphyrin lies the distal haem pocket where the accommodation of the gaseous ligand takes place. It is the folding of the globin which supplies selective reversible binding, while the iron covalently binds the ligand[3], stabilised by distal HisE7 of the globin[4].

HISTORY OF HAEMOGLOBIN

Scientific interest toward mammalian Haemoglobin started two centuries ago, yet its structure was resolved only in 1959 by Max Perutz, who was assigned a Nobel prize three years later (**Figure 2**). The size of research initiated by Haemoglobin curiosity is considered to have originated molecular medicine and promoted its advancement.

HUMAN HAEMOGLOBIN

The erythrocytes' Haemoglobins found in the human organism are built by the symmetric pairing of a heterodimer of α and β globin, into a tetrameric unit with noteworthy biological roles, including distribution of Oxygen from lungs to the rest of the body and specific interactions with carbon dioxide, carbon monoxide and nitric monoxide under distinctive conditions. Subunits α of 141 residues and β respectively fold into 7 and 8 α helices, consolidating the molecule in globular conformation and bringing the proximal and distal histidine residues in the obligatory arrangement to permit the binding and transport of O_a, CO or NO. Discrimination of ligands is due to the steric hindrance imposed by the geometry of the bond of haem iron with ligand[3]. On the other hand, the transport of CO2 in blood solution does not involve binding to haem central iron. Nevertheless, the interaction of the gaseous ligand with the amino-terminal residues of haemoglobin as a weak carbamino complex is required[2].
Throughout an individual's lifespan, three different types of Haemoglobin dominate in abundance during specific developmental phases, Embryonic Haemoglobin, Fetal Haemoglobin and Adult Haemoglobin HbA or HbA2, found in significantly smaller quantities. The dissimilarity among these three Haemoglobin variants stands in the type of α and β chain of which they are composed. The variousness in tetramer composition leads to the diverse subunit interface stability hence changes in Oxygen affinity and binding[5]. Genes that encode alternative α and β chain subunits are correspondingly localized in a gene cluster in chromosome 16 and chromosome 11 in mammals, deemed to be separated during evolution(**Figure 1**). Must be emphasized the sequential pattern of activation and transcription of these gene clusters, especially the β chain one: passing from Embryo to an adult human, activation is shifted from 5' to 3' of the gene clusters. Regardless, in the erythrocytes of normal human adults, 97% of haemoglobin is HbA (α2β2), 2% is Haemoglobin A2 (α2δ2) and 1% is fetal (α2γ2)(**Figure 1b)** proving that the regulation of subunits expression is not a strict ON/OFF switch[2].
The model proposed by J.Monod, J.Wyman and J.P.Changeux[6] suggests that the Hb adopts at least two quaternary structures R for relaxed and T for tense states, which differ due to movement of α1β1 and α2β2 dimmers. It causes altering of α1β2 interactions interface and structural changes that include the haem and contiguous structural elements like HisF8[3]. Concerted changes within the haem pockets imply different ligand affinities of each state. T state deoxyhaemoglobin has a lower affinity of Oxygen binding than R state oxyhaemoglobin. What several studies have concluded is the sequential affinity of Haemoglobin for oxygen. The binding of an Oxygen molecule in the T state Haemoglobin, do not directly switch to the R state Haemoglobin with increased affinity for oxygen but induces conformational rearrangements of tertiary and quaternary structure that favour another oxygen molecule binding to Haemoglobin, increasing the probability of bringing the molecule to the R state. Hence the more oxygen binds to a molecule of Hb, the more its affinity for O_2 increases[3].

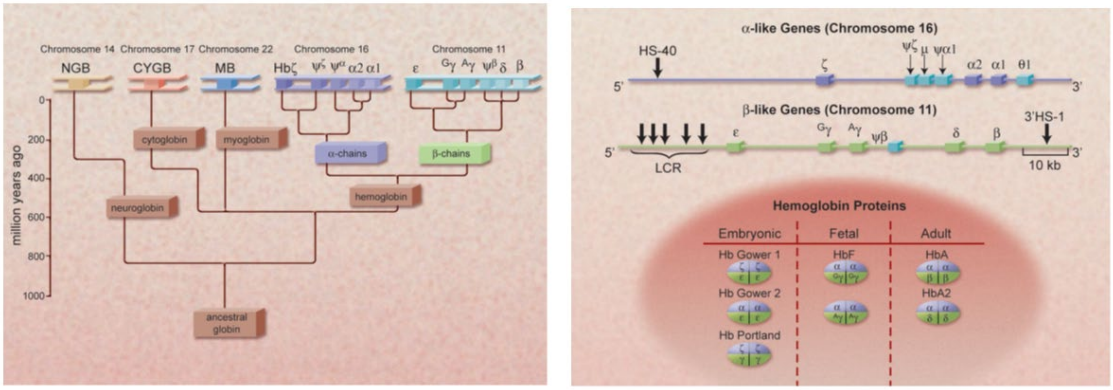


Figure 1. a)*The proposed the evolutionary relationship of human globin based on sequence similarities..* ***b)****The structure of α-like and β-like gene clusters in human genome, together with different Haemoglobins composition. Figures taken from article 3.*

METHODS

To accomplish this study, wild type human HbA Haemoglobin structure in oxy T state with all haems occupied by Oxygen and with no mutation detected is retrieved from RCSB PDB. The structure with PDB identifier 1GZX is acquired by X-Ray Crystallography, with a 2.10Å resolution and acceptable quality values for all validation parameters.

The whole content of IntAct database is downloaded and then filtered in order to collect only data regarding human interactome. The download, management and preliminary filtering and modification of files are executed in Linux terminal. Afterwards Python script is written to parse the files information and perform the required computations to achieve the results.

PROBLEM 1:

To perform the search and analysis of the interactions within and between monomers and with the haem groups, the PDB file of 1GZX has been downloaded from RCSB PDB website[7], containing atomic coordinates of the studied structure. In the 1gzx.pdb file, a wealth of information is provided, including authors, literature, reference and methods to achieve structure determination, as well as informative records regarding atom and residue organization of the protein. The focus of our analysis are the lines starting with ATOM, which refers to atoms of the haemoglobin residues, and HETATM, which includes lines with coordinates of atoms belonging to nonstandard residues, like inhibitors, ions, solvent, prosthetic group etc. In this case, HETATM includes coordinates of atoms from haem group and water. Exploiting the information regarding the .pdb[8] file organization, in a Python script, the 1gzx.pdb file information is accessed to select the 'ATOM' lines with coordinates of peptide atoms and 'HETATM' lines with coordinates of atoms belonging to haem groups, to be organized in 'atom' and 'hem' dictionary respectively. This procedure is applied to every chain of Haemoglobin. For every atom in pdb file, belonging to either standard or nonstandard residue, among the data regarding the residue and peptide chain to which it belongs there are furthermore given the coordinates of this atom in X, Y and Z plans.

To our analysis, these three specific data columns are crucial to determine the distances between two atoms and identify a possible type of interaction that can exist between them. A basic function that applies the mathematical formula of the distance between two points is used to compute the distance between:

- Every peptide atom, to haem and haem-bound oxygen atom. Every pair with distance 3.5 Å and lower is saved into a file to undergo further manual filtering to detect possible H bonds formed between the monomer and its oxy haem group.
- Atoms belonging to two different chains of Haemoglobin. Every pair of atoms with a distance 4 Å and lower can form a potential salt bridge, participating in the interaction interface between two chains. These criteria-fulfilling pairs are saved into a file, to get manually filtered considering the type of atoms and residues which can chemically interact through a salt bridge with each other.

Additional functions have been added to the script to create supplementary files in which interactions with the haem group and other chains are reported based on the implicated residues.

For accuracy in the selection of probable Hydrogen bonds and salt bridges interactions, the structure is visualized in CHIMERA[9], to check the positioning of the involved residues.

PROBLEM 2:

PDB files for each monomer and all possible trimers of haemoglobin chains have been generated using commands of Linux terminal and together with 1gzx.pdb file are utilised to create DSSP[1] entries. This is performed by DSSP[10,11] program which given the full and valid 3D structure of a protein computes the most likely secondary structure assignment of it, extracting information from the coordinates. Additionally, it provides information regarding the geometrical features, solvent exposure, dihedral angles, etc. By a Python script that parses the dssp files, information regarding residues is organized into a dictionary, including amino acid symbol and identifier, secondary structure in which the residue is involved, solvent accessibility, dihedral angles, as well as manually computed Relative solvent accessibility based on residues accessibility values[2] by B.Rost and C. Sander[12].

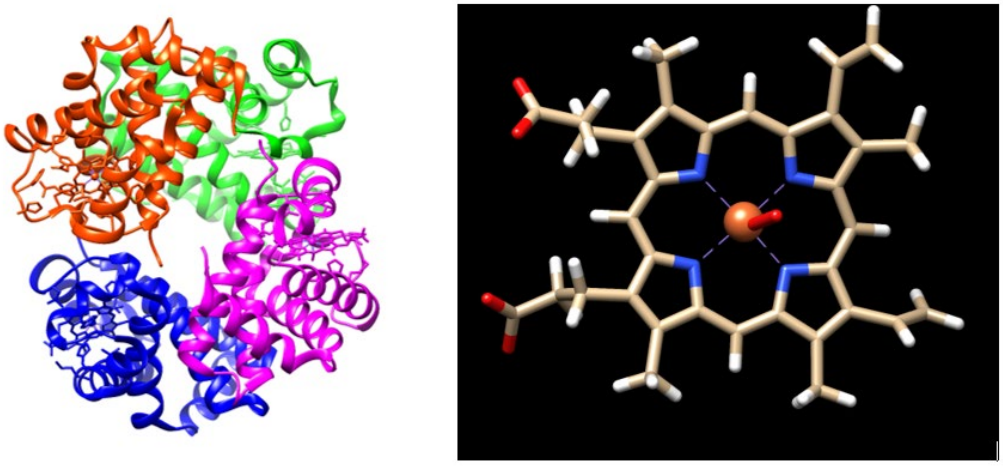


Figure 2 a) *Human Adult HbA Haemoglobin, coloured in orange chain A, in green chain B, in magenta chain C and in blue chain D. Helices and loops representing the polypeptide sequence, while Haem group left in stick and ball representation.* ***b)*** *Structure of haem group / Protoporphyrin IX, in orange represented the central Iron and a molecule of Oxygen (ligand) represented in red.*

[1] DSSP, is a database of secondary structure assignment for all protein entries in pdb

[2] measure of residue solvent exposure

An explicit function in Python script computes the solvent accessibility for a given chain/monomer, by summing up the solvent accessibility of its every residue. To understand the solvent accessibility loss of chain A due to interactions between pairs of monomers, taken for example A and B, the overall solvent accessibility of A in tetramer and in the trimer that lacks monomer B, are compared. Their numerical difference results in the loss of solvent accessibility area of A when bound to B, which is as well a measure of the interaction strength between two monomers.

For each residue of a given monomer, are computed solvent accessibility and Relative solvent accessibility, respectively by comparing solvent accessibility values and relative solvent accessibility of a residue in monomer and tetramer form, to understand the main residues in interaction hotspots between monomers. The numerical differences are saved in text files.

PROBLEM 3:

The file retrieved from IntAct[13] Molecular Interaction Database (Footnote, an open-source database system and analysis tool providing molecular interaction data) enclosed all possible interactions documented in this database. Consequently, the selection of interactions of interest was required, to adapt the data to the purpose of our study.

The filtering criteria included:

1) Type of interaction: Direct interaction

2) to be considered only Human proteins having their UniprotID [14]

3)Isoforms to be converted to native forms to lower the complexity of the network

This has been performed on the Linux command line, while the rest of the network analysis was made on Python script. File parsing follows the same practice as in the previous study steps, subtracting information regarding interactions and organising it in two dictionaries, one for the network nodes and the other for the edges. For every pair of protein-protein interactions, two interacting proteins are added in the dictionary of nodes and an undirected edge of format (protein1, protein2) is added in the dictionary which keeps track of edges. These nodes and edges-containing data structures are utilized by the NetworkX[15] package of Python, which allows creation, manipulation, visualization and exhaustive study of networks. The analysis of the structure, including information regarding the degree of connectivity, clustering, betweenness, and manipulation of the Haemoglobin interaction network, have been conducted employing the built-in functions of the NetworkX[15] package.

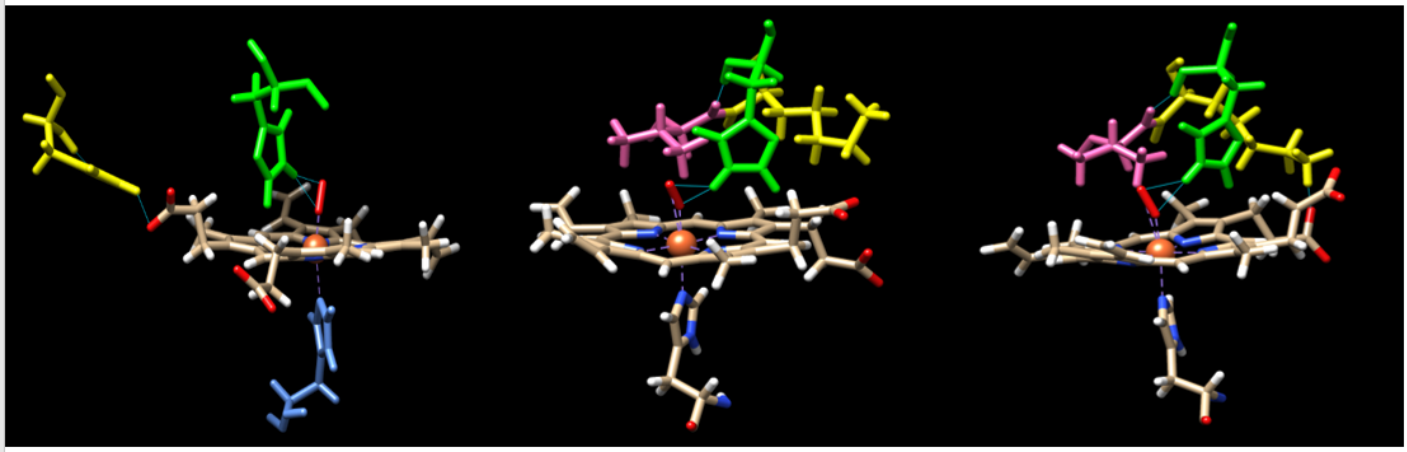


Figure 3 *A comparison between haem pocket of Chains A, B and D.* ***In the first structure****, Haem Pocket of Chain A, Proximal Histidine His87 coloured in blue and in pale purple lines its covalent interaction with Central Iron of haem is depicted. His45 which interacts with propionate tail through an H bond coloured in yellow and the distal Histidine His58 which interacts with Oxygen though two H bonds coloured in green.* ***The second and third structure*** *colouring differs, the proximal Histidine is left in atom-based colouring, distal histidine remains green, Val210 and Val610 depicted in pink and Lys209 and Lys609 in yellow. In the second structure, which corresponds to pocket of chain B, the orientation of Lys209 does not enable it to bind to haem group, while the third structure, representing the pocket of chain, shows a H bond formed between the yellow Lys609 and one tail of propionate (H bond signed in cyan). Moreover the Val210 and Val610 orientation towards the Haem and the distal Histidine differs from chain B to chain D, a difference that might be related to T and R conformation switches, and/or affinity and mechanism of Oxygen binding.*

Chain	Residue	Hetero-Residue	Atoms in distance ≤3.5Å	Bond type
A	HIS45	HEM1142	NE2-O2D	Hydrogen bond
C	HIS445	HEM1542	NE2-O2D	Hydrogen bond
D	LYS609	HEM1690	NZ-O1A	Hydrogen bond
A	HIS58	OXY1143	NE2-O1, NE2-O2	Hydrogen bond
B	HIS206	OXY1291	NE2-O1, NE2-O2	Hydrogen bond
C	HIS458	OXY1543	NE2-O1, NE2-O2	Hydrogen bond
D	HIS606	OXY1691	NE2-O1, NE2-O2	Hydrogen bond

Table 1 *Atoms of every chain and Haem/Oxygen in adjacency up to 3.5Å that interact through H Bonds, to mediate the binding of the Haem with the polypeptide.*

Chain	Residue	Hetero-Residue	Atoms Interacting	Bond type
A	HIS87	HEM1142	NE2-FE	Covalent bond
B	HIS235	HEM1290	NE2-FE	Covalent bond
C	HIS487	HEM1542	NE2-FE	Covalent bond
D	HIS635	HEM1690	NE2-FE	Covalent bond

Table 2 *Atoms of every chain which covalently bind to the central iron of the Haem group, stabilizing it and the haem pocket.*

RESULTS

PROBLEM 1:

The goal of the first part of this study is the observation of important interactions among the monomer residues with the haem group and the Oxygen. Key idea is to check the adjacency of residues with the prosthetic group and its ligand due to peptide folding, which provides the means for stable interaction and anchoring of them in the heart of the monomer. Distances between residues and haem and oxy group are computed, and only those less than 3.5 Å, which corresponds to the maximal distance favouring a Hydrogen Bond, are considered in the further step of the examination. (Table 1 of Supplementary material). For the formation of a Hydrogen bond, an electron donor and an acceptor should be involved in the interaction. Therefore, the results of distance computation in Python are checked whether the atoms involved belong to any of the aforementioned categories, such that an interaction can be encouraged. (**Table 1**)

For chains A and C, one histidine, His45, forms a Hydrogen bond with the haem propionate tail, deemed to be one of the crucial

interactions for the prosthetic group to be kept bound to the α subunit peptide. Surprisingly, subunits B show a slightly different pattern, chain B in T oxy state does not form any Hydrogen bond with the haem group, while chain D involves Lysine 609 in the Hydrogen bonding, rather than a Histidine. (**Figure 3**) This finding is consistent with conclusions of the "Crystal structure of T state haemoglobin with oxygen bound at all four haems"[16] where different conformational changes are noticed to happen in each chain because of Oxygen binding. Despite sharing an identical peptide sequence, the ligand binding dynamics of chains B and D induce a diverse configuration of its haem pocket.

Two Hydrogen bonds are formed between each atom of ligand and the distal Histidine of each chain, specifically with the NE2 of its imidazole ring, of high impact in the stabilization of Oxygen in the haem pocket. Additional contribution in Oxygen binding is delivered by the Oxygen – Central iron polar bond[3]. (**Table 2**)

A conserved interaction of chain residues is with the haem group consisting of the covalent bond between a lone pair of NE2 in imidazole of proximal Histidine (HisF8) and central iron atom. The additional role of this Histidine includes mediating iron reactivity, through Hydrogen bonding with other residues of the haem pocket[3].

Valine is another important factor in the formation and functionality of the haem pocket (**Table 3**). Numerous Valine substitution studies[17] have concluded that its shape, size and polarity affect the conformation of the distal pocket and the stability of iron-ligand. The effect of Valine in the activity of the haem group is the same in both subunits, however, its interaction in subunit α may involve both haem and oxygen, while in β interaction might occur with the haem group only. These assumptions are done considering the adjacency of Valine's atoms to those of haem and oxygen.

A list of possible interactions between chains is generated, with the interatomic distance of residues smaller than 4A, such a way that Hydrogen bonds together with Salt bridges can be detected.(Table 2 of Supplementary Materials). Selecting interactions between opposite charge-residues, only 4 salt bridges appear to be contributing to the tetramer formation, all involving chain A. (**Table 4**). Surprisingly there is not a conserved pattern of salt bridge interaction between α and β subunits or in the connection of chains of the same subunit. Thus, salt bridges cannot be considered a major source in the interaction interface between monomers. Several Hydrogen bonds keep the tetramer together, some extending from the subunit of one heterodimer to the other. Identical chains A and C symmetrically interact with each other through a Lysine and an Arginine, however, chains B and D do not have any contact or adjacent residues to each other. (**Table 5**). Chemical conditions for possible Hydrogen bonds formation between chains A and D and chains B and C exist, interactions in both cases involve the same type of residues making A-D and B-C interaction patterns identical.

Chain	Residue	Hetero-Residue	Atoms in distance ≤3.5Å
A	VAL62	OXY1143	CG2-O2
A	VAL62	HEM1142	CG1-CAC
B	VAL210	OXY1291	CG2-O1 , CG2-O2
C	VAL462	OXY1543	CG2-O2
C	VAL493	HEM1542	CG1-CAC
D	VAL610	OXY1691	CG2-O1, CG2-O1

Table 3 *List of Valine residues found in Haem pocket of every chain, which have important functional and structural roles in the haem group of protein and especially related to Oxygen affinity and binding stability. The distances of some Valine atoms to the haem and oxygen ones are reported.*

Chain	Residue	Chain2	Residue2	Atoms in distance ≤ 4Å
A	GLU23	B	LYS263	OE2-NZ
A	ASP126	C	ARG541	OD2-NH2, OD2-NH1
A	ARG141	C	ASP526	NH2-OD2, NH1-OD2
A	ARG92	D	GLU586	NH2-OE1, NH1-OE1

Table 4 *List of Salt bridges created between monomers that aid the formation of the tetramer. The involved residues are reported, together with the interacting atoms.*

Chain	Residue	Chain2	Residue2	Atoms in distance ≤ 3.5Å
A	ARG31	B	PHE265	NH1-O, NH2-O
A	ARG31	B	THR266	NH1-O
A	ARG31	B	GLN270	NH1-OE1
A	HIS103	B	GLN274	NE2-OE1
A	PRO114	B	HIS259	O-NE2
A	PHE117	B	HIS259	O-NE2
A	PHE117	B	ARG173	O-NH1, O-NH2
A	LYS127	C	ARG541	NZ-O, NZ-OXT
A	ARG141	C	LYS527	O-NZ, OXT-NZ
A	LYS40	D	HIS689	NZ-OXT
A	TYR42	D	ASP642	OH-OD1
A	LEU91	D	ARG583	O-NH2
A	ASP94	D	TRP580	OD1-NE1
A	ARG141	D	VAL577	NE-O, NH1-O
B	HIS289	C	LYS440	OXT-NZ
B	VAL177	C	ARG541	O-NH1, O-NE
B	TRP180	C	ASP494	NE1-OD1
B	ARG183	C	LEU491	NH2-O
B	ASP242	C	TYR442	OD1-OH
C	PRO514	D	HIS659	O-NE2
C	PHE517	D	ARG573	O-NH1, O-NH2
C	PHE517	D	HIS659	O-NE2
C	HIS522	D	ARG573	ND1-NH2
C	ARG431	D	PHE665	NH2-O, NH1-O
C	ARG431	D	GLN670	NH1-OE1
C	SER435	D	GLN670	OG-NE2
C	HIS503	D	GLN674	NE2-OE1

Table 5 *All the Hydrogen bonds formed between monomers which possibly play an important role in the formation of tetramer.*

PROBLEM 2:

Solvent accessibility of a protein's residues provides scientists with important insights regarding protein fold, interaction strength and stability. Loss of Solvent Accessibility is computed to evaluate the surface of interaction between every pair of chains. Interaction between α and β subunits of a heterodimer[3], appear to be the strongest in the tetramer. Although chain A and chain B are sequentially identical to chain C and chain D, C-D have a stronger interaction between them, numerically involving even a larger interaction interface. (**Table 6**). Particular is the evidence of almost total lack of interaction between chain B and D, verified by a low value of Loss of Solvent Accessibility, meaning that the Solvent Accessibility of chain B is barely affected, whether chain D is or is not bound to ABC trimer. Analysing the Relative Solvent Accessibility of each chain's residues of Haemoglobin in monomeric, when the peptide is not interacting with any other subunit, and in the tetrameric state, when the tetrameric protein is formed, aids in recognizing some important residues involved in subunits interaction interface. (Table 3 of Supplementary Materials). Around 150 amino acids experience a change bigger than 10% in their RSA from monomeric to tetrameric form, indicating their involvement in interaction surface and specific folding which ensures tetramer formation. Some of them, especially hydrophobic[4] residues, become fully inaccessible in tetrameric protein, proving that the tetrameric form also stabilizes thermodynamically the peptide chains. (**Table 7**).

[3] A heterodimer is formed by one alpha and one beta subunit, like A-B and C-D

[4] Determination of hydrophobic residues by hydrophobicity scale of Kyte & Doulittle[18]

Chain not bound to trimer/ Chain of trimer	A	B	C	D
A	/	858	241	697
B	886	/	659	24
C	238	670	/	809
D	671	22	826	/

Table 6 *The change in solvent accessibility of a chain, when it interacts with another, is shown in the table for every pair of monomers. Take for example chain A and B. Chain A is bound to C and D forming a trimer, but does not interact with chain B yet, each monomer of trimer is characterized by a Solvent accessibility. When chain B joins the trimer and starts interaction with its components, there is a loss of solvent accessibility for all chains A, C and D due to their newly established interactions with B. The loss for A is 858 Å², for C 670 Å², and for D 22 Å².*

Chain	Residue	Monomer RSA	Tetramer RSA	Loss	Loss in %
D	VAL81	0.5704	0	0.5704	100
B	GLN81	0.4091	0	0.4091	100
D	GLN80	0.404	0	0.404	100
B	ALA39	0.3679	0	0.3679	100
C	VAL52	0.3662	0	0.3662	100
D	ALA38	0.3585	0	0.3585	100
A	VAL50	0.3521	0	0.3521	100
A	ALA35	0.3302	0	0.3302	100
C	ALA35	0.3302	0	0.3302	100
D	PHE22	0.1117	0	0.1117	100
B	PHE18	0.0914	0	0.0914	100
D	VAL9	0.0634	0	0.0634	100
A	CYS3	0.0222	0	0.0222	100
C	CYS3	0.0222	0	0.0222	100

Table 7 *Some Hydrophobic residues with major Relative Solvent Accessibility loss due to monomer-monomer interactions are reported. Due to their 100% RSA loss, we can surely state that they are found in hotspots of inter-monomeric interactions.*

PROBLEM 3:

The whole network of human direct protein-protein interaction subtracted by IntAct database content is considered to depict the role and connectivity of α and β subunits in it. Application of built-in function of NetworkX to construct and analyse the structure resulted in a total network with 4257 human proteins (**Figure 4)**, involved in 7067 direct interactions, with 550 connected components and the major component that includes both α and β subunits. A and β in the network show the parameters as in the following figures, compared to the average values of all participating proteins. (**Figure 5)**.

To be emphasized is the 0 clustering value for both α and β nodes. To ascertain it, the subnetwork with the neighbours of α and β nodes is visualized and it is clear that nor the neighbours of α subunit, nor those of β are interconnected with each other, hence the clustering corresponds to zero.

With the removal of node representing subunit α of Haemoglobin (table), the degree of β is 1, making β node a dead-end. Therefore, there is not any short path of two nodes of the network that passes through β node, and its betweenness becomes zero.

In a network without β node, its clustering remains 0 and degree lowers to 6. Betweenness goes from 24050.55 to 20800, since several shortest paths crossing α node also pass through the node of β subunit. This change in value of betweenness emphasizes the connectivity of α and β even in interactome.

Modifying the network by cancelling the edge, the interaction between α and β subunit, there is one short path of three edges which connects α and β nodes.

'P68871' - 'P02008' - 'P02100' - 'P69905'(**Figure 6)**

Since the differences in network parameters are inconsiderable, we can assume that subunit α and β are involved in direct interaction, however the role that they play in the overall human interactome is not significant.

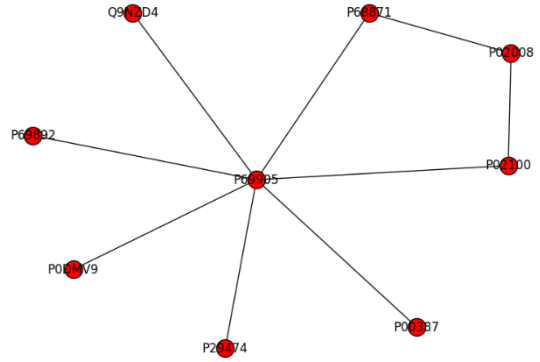
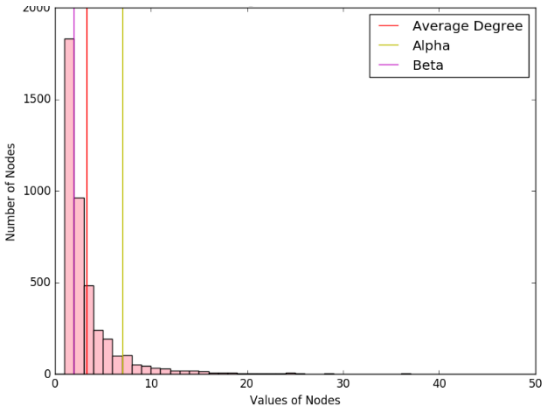
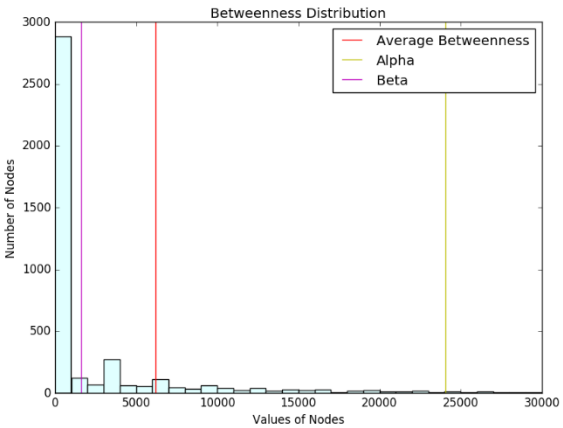


Figure 4 *Alpha And Beta subunits together with their neighbours subgraph.*





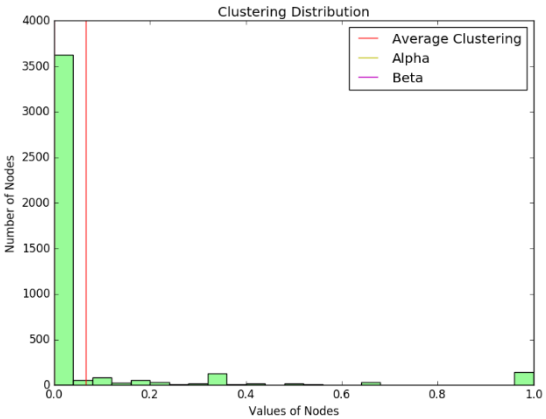


Figure 5 *Representation of Degree (in pink), Cluster Coefficient (in green) and Betweenness (in blue) of Human Direct Protein-Protein Interaction Network through histograms, pointing the mean value of the network nodes and values of our nodes of interest α and β, in order to compare their connectivity, centrality and importance in the network.*

To notice is the clustering coefficient of both alpha and beta node which scores 0.

CONCLUSION

60 years of structural and functional study of Haemoglobin have expanded horizons in scientific research by allowing allosteric theory to develop and revolution of methods like protein crystallography. Nonetheless, these decades have not been sufficient to entirely comprehend Haemoglobin biomechanics, in vivo unpredictable behaviour and complex folding dynamics.

Haemoglobin is formed by symmetrical repetition of the α-β heterodimer. Although heterodimers are identical regarding residue sequence, their subunits interaction is of different intensity with chain C and chain D interacting stronger, haem pocket compositions changes, as well as the monomer-monomer interactions the subunits are involved in with two α subunits able to interact with each other, regardless, this does not occur between β subunits.

- There are two Histidine residues per haem pocket, HisF8 and HisE7, which respectively enable the covalent bond with haem group and binding and stabilization of ligand. Of high interest is the choice of chain D, where Lys609 together with His606, stabilize two propionate tails. It is not known why the arrangement of residues in the haem pocket differs from chain B, which shares the same polypeptide sequence with chain D. The various conformational changes haem pockets of different chains experience upon oxygen binding, might be a consequence of this Lysine selection and vice versa.
- Salt bridges are not the dominating electrostatic interaction that keeps the tetramer together, instead, the Hydrogen bonds appear to be more frequent in the interaction interface between monomers.
- The binding of chains all together is mediated by diverse residues, and it aids the monomers to reach the most stable state, hiding their hydrophobic residues in the interaction interface, mainly Valine, Alanine, Cysteine and Proline.
- Although Haemoglobin is truly vital to human organisms, its subunits are not central components in the human direct interaction network, classified through the least interacting and connected nodes of it. Their interactions mainly include alternative subunits of Haemoglobin transcribed from α and β gene clusters, chaperon and protein stabilizers like A Haemoglobin Stabilizing Protein. **(Table 8) (Table 9).**

Haemoglobin is still accompanied by question marks nowadays; plenty is left for the new researcher generation to answer. Answers that would unveil not only a piece of the complicated puzzle of human biology but also encourage innovation in therapies and clinical research, bearing hope.

Protein Uniport ID	Protein interacting with β
P02008	Hemoglobin subunit zeta
P02100	Hemoglobin subunit epsilon
P69905	Hemoglobin subunit alpha

Table 8 *Proteins of human organism that directly interact with subunit β or Haemoglobin*

Protein Uniport ID	Protein interactin with α
P68871	Hemoglobin subunit beta
P00387	NADH-cytochrome b5 reductase 3
P29474	Nitric oxide synthase, endothelial
P0DMV9	Heat shock 70 kDa protein 1B
P69892	Hemoglobin subunit gamma-2
Q9NDZD4	Alpha-hemoglobin-stabilizing protein

Table 9 *Proteins of human organism that directly interact with subunit α or Haemoglobin*

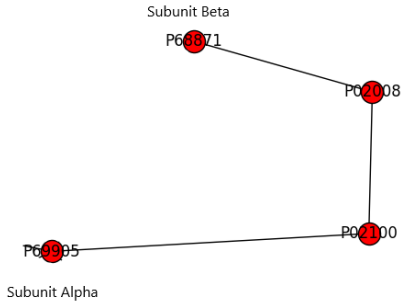


Figure 6 *Shortest Path from Subunit Alpha to Beta, if the edge connecting them is removed from the network.*

REFERENCES

1. Tobi, D. Dynamics based clustering of globin family members. *PloS One* **13**, e0208465 (2018).
2. Schechter, A. N. Hemoglobin research and the origins of molecular medicine. *Blood* **112**, 3927–3938 (2008).
3. Gell, D. A. Structure and function of haemoglobins. *Blood Cells. Mol. Dis.* **70**, 13–42 (2018).
4. Perutz, M. F. Myoglobin and haemoglobin: role of distal residues in reactions with haem ligands. *Trends Biochem. Sci.* **14**, 42–44 (1989).
5. Manning, L. R. *et al.* Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. *Protein Sci. Publ. Protein Soc.* **16**, 1641–1658 (2007).
6. Monod, J., Wyman, J. & Changeux, J. P. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. *J. Mol. Biol.* **12**, 88–118 (1965).
7. Bank, R. P. D. RCSB PDB: Homepage. https://www.rcsb.org/.
8. PDB101: Learn: Guide to Understanding PDB Data: Introduction. *RCSB: PDB-101* https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction.
9. UCSF Chimera Home Page. https://www.cgl.ucsf.edu/chimera/.
10. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411-419 (2011).
11. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
12. Rost, B. & Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins* **20**, 216–226 (1994).
13. Orchard, S. *et al.* The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2014).
14. UniProt. https://www.uniprot.org/.
15. Search — NetworkX 2.6.2 documentation. https://networkx.org/documentation/stable/search.html?q=.edges.
16. Paoli, M., Liddington, R., Tame, J., Wilkinson, A. & Dodson, G. Crystal structure of T state haemoglobin with oxygen bound at all four haems. *J. Mol. Biol.* **256**, 775–792 (1996).
17. Quillin, M. L. *et al.* Structural and functional effects of apolar mutations of the distal valine in myoglobin. *J. Mol. Biol.* **245**, 416–436 (1995).
18. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).