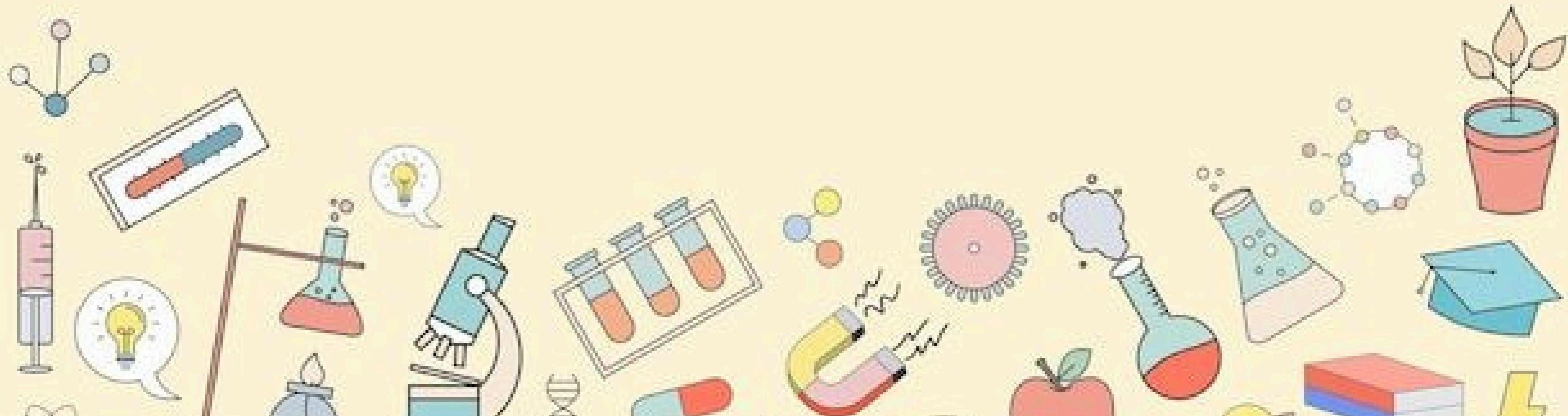
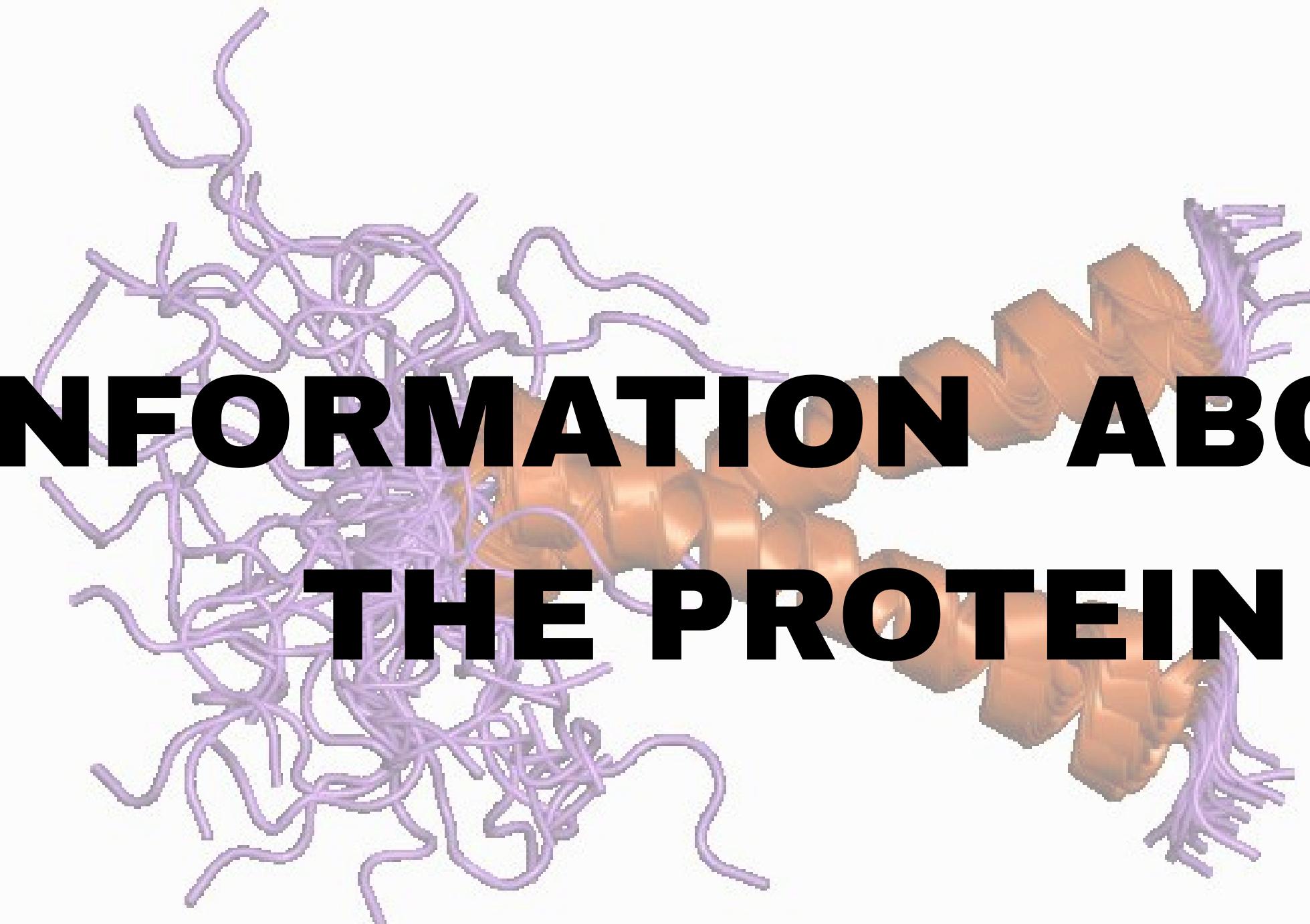


STRUCTURAL AND FUNCTIONAL ANALYSIS OF GLYCOPHORIN A





INFORMATION ABOUT THE PROTEIN

(Data Retrieved from UniProt)

NAME OF THE PROTEIN: Glycophorin-A

ORGANISM: *Homo Sapiens* (Human)

PROTEIN SEQUENCE LENGTH: 150 Amino acids

GENE: GYPA

PRIMARY ACCESSION CODE: P02724

The screenshot shows the UniProt protein details page for P02724 · GLPA_HUMAN. The top navigation bar includes links for BLAST, Align, Peptide search, ID mapping, SPARQL, UniProtKB, Advanced, List, Search, and Help. The main content area displays the following information:

Function	P02724 · GLPA_HUMAN	
Names & Taxonomy	Protein ⁱ	Glycophorin-A
Subcellular Location	Gene ⁱ	GYPA
Disease & Variants	Status ⁱ	UniProtKB reviewed (Swiss-Prot)
PTM/Processing	Organism ⁱ	Homo sapiens (Human)
	Amino acids	150 (go to sequence)
	Protein existence ⁱ	Evidence at protein level
	Annotation score ⁱ	5/5

PROTEIN SEQUENCE IN FASTA FORMAT:

>SP|P02724|GLPA_HUMAN GLYCOPHORIN-A OS=HOMO SAPIENS OX=9606 GN=GYPA PE=1 SV=3
MYGKIIFVLLLSEIVSISALSTTEVAMHTSTSSSVTKSYISSQTNDTHKRDTYAATPRAHEVSEIS
VRTVYPPEEETGERVQLAHHFSEPEITLIIFGVMAVGIGTILLISYGIRRLIKKSPSDVKPLPSPD
TDVPLSSVEIENPETSDQ

SUBCELLULAR LOCATION: RBC membrane

Single Pass Type I Membrane Protein

Appears to be colocalized with SLC4A1

DOMAIN: It is a single domain protein with a domain length of 34-146 (112 amino acid long residue).

P02724 Glycophorin-A
UniProtKB/Swiss-Prot protein

This protein matches this entry:

1 - 1 of 1 entry matching Pfam

Overview Entries 1 Structures 16 AlphaFold 1 Sequence Similar Proteins 326

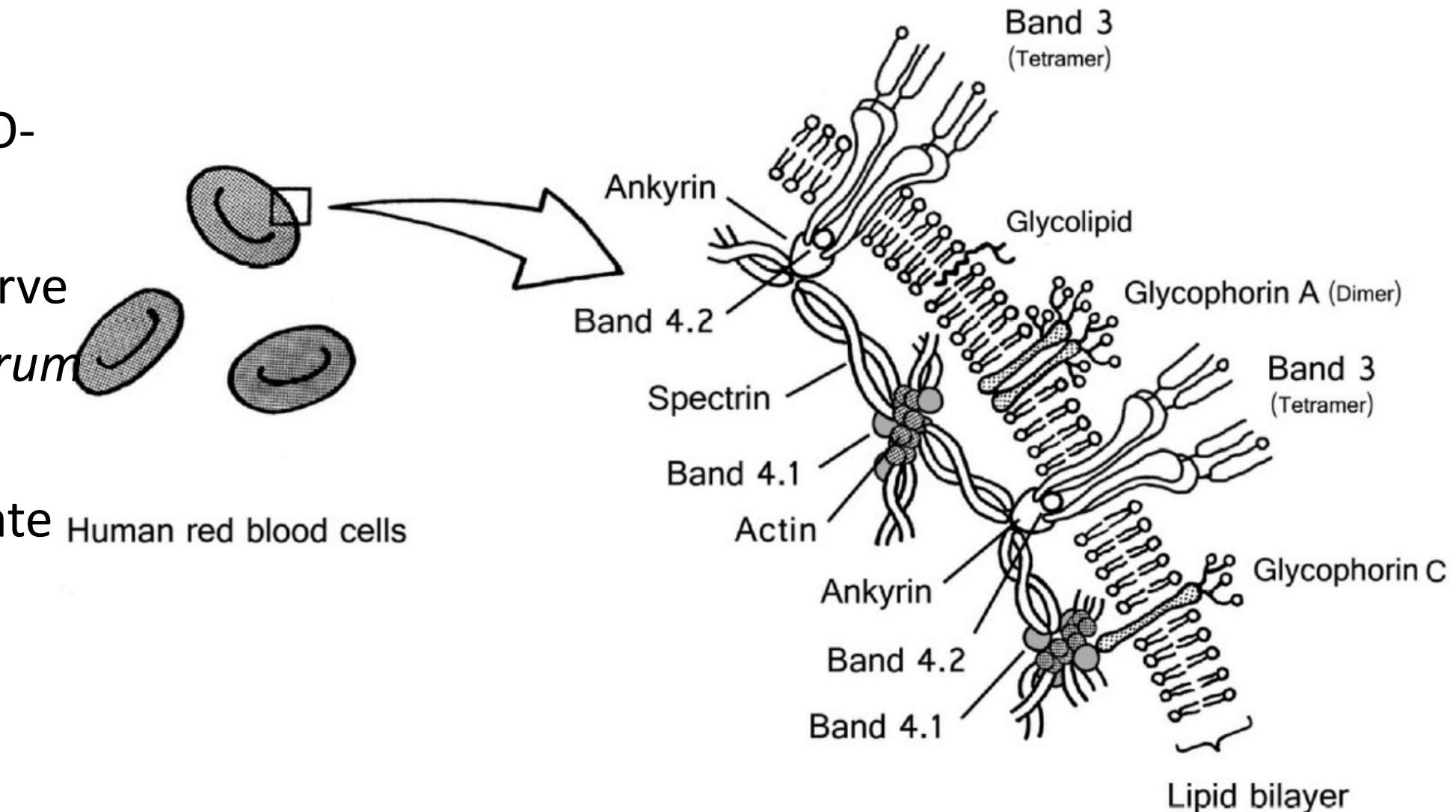
Accession	Short Name	Name	Source Database	Matches
PF01102	Glycophorin_A	Glycophorin A	pfam - PF01102	34-146

Search Download

FAMILY: Belongs to Glycophorin-A family.

BINDING SITES:

1. Glycophorin-A's N-terminal is heavily glycosylated with O-linked glycans and one N-linked glycan.
2. These sugar chains contain sialic acid residues, which serve as binding sites for malaria parasite *Plasmodium falciparum*.
3. Glycophorin-A also provides a binding site for influenza virus, recognizing the neuraminic acid on its carbohydrate chains.



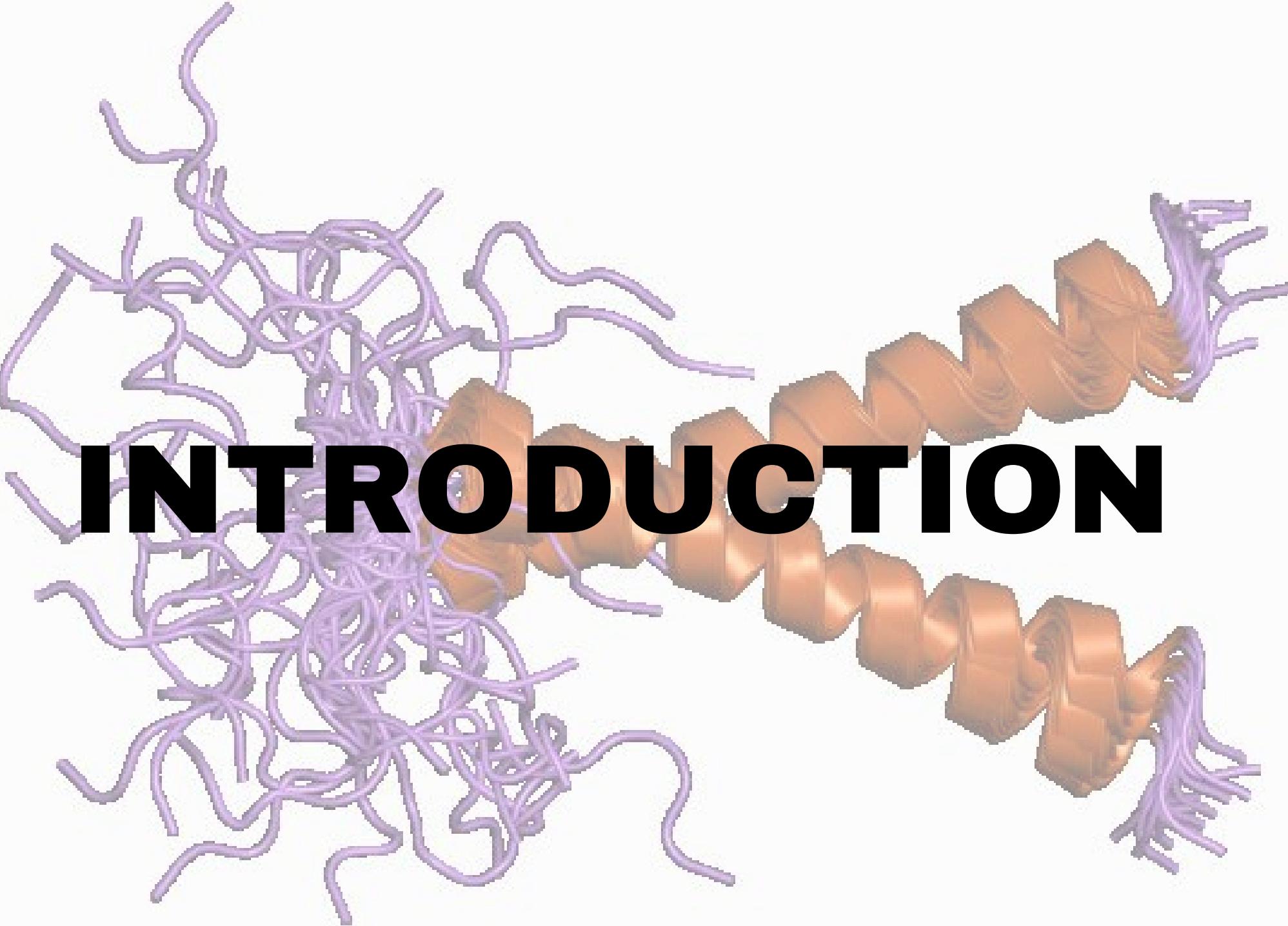
MAJOR FUNCTIONS:

Glycophorin A (GPA) is a type I single-pass membrane sialoglycoprotein essential for structural stability, cell interactions, and receptor-mediated functions.

- **Membrane stability:** Maintains the RBC shape, stabilizes the membrane through Band 3 association.
- **Membrane Organization:** Heavily O- and N-glycosylated, GPA provides a strong negative surface charge, preventing RBC aggregation and unwanted adhesion.
- **Protein Folding & Trafficking:** Acts as a chaperone for Band 3, aiding its folding and membrane localization, ensuring efficient anion exchange and ionic balance.
- **Antigen Presentation:** Carries M and N blood group antigens and forms the Wr^b complex with Band 3, influencing transfusion compatibility.
- **Pathogen Receptor:** Serves as a receptor for Plasmodium falciparum, influenza, reoviruses, and certain E. coli strains via sialic acid residues.
- Appears to be important for the function of SLC4A1 and is required for high activity of SLC4A1. May be involved in translocation of SLC4A1 to the plasma membrane.

POST TRANSLATIONAL MODIFICATION:

Glycophorin-A undergoes extensive **O-linked glycosylation**, with the major glycans being NeuAc- α (2-3)-Gal- β (1-3)-[NeuAc- α (2-6)]-GalNAcOH (~78%) and NeuAc- α (2-3)-Gal- β (1-3)-GalNAcOH (~17%). Minor forms (~5%) include disialylated structures such as NeuAc- α (2-8)-NeuAc- α (2-3)-Gal- β (1-3)-GalNAcOH. Approximately 1% of all O-linked glycans carry blood group A, B, or H determinants, derived from a type-2 precursor core (Gal- β (1-3)-GlcNAc- β 1-R). These antigens form through the sequential addition of fucose (H antigen) and then N-acetylgalactosamine or galactose for A and B antigen specificity, respectively.



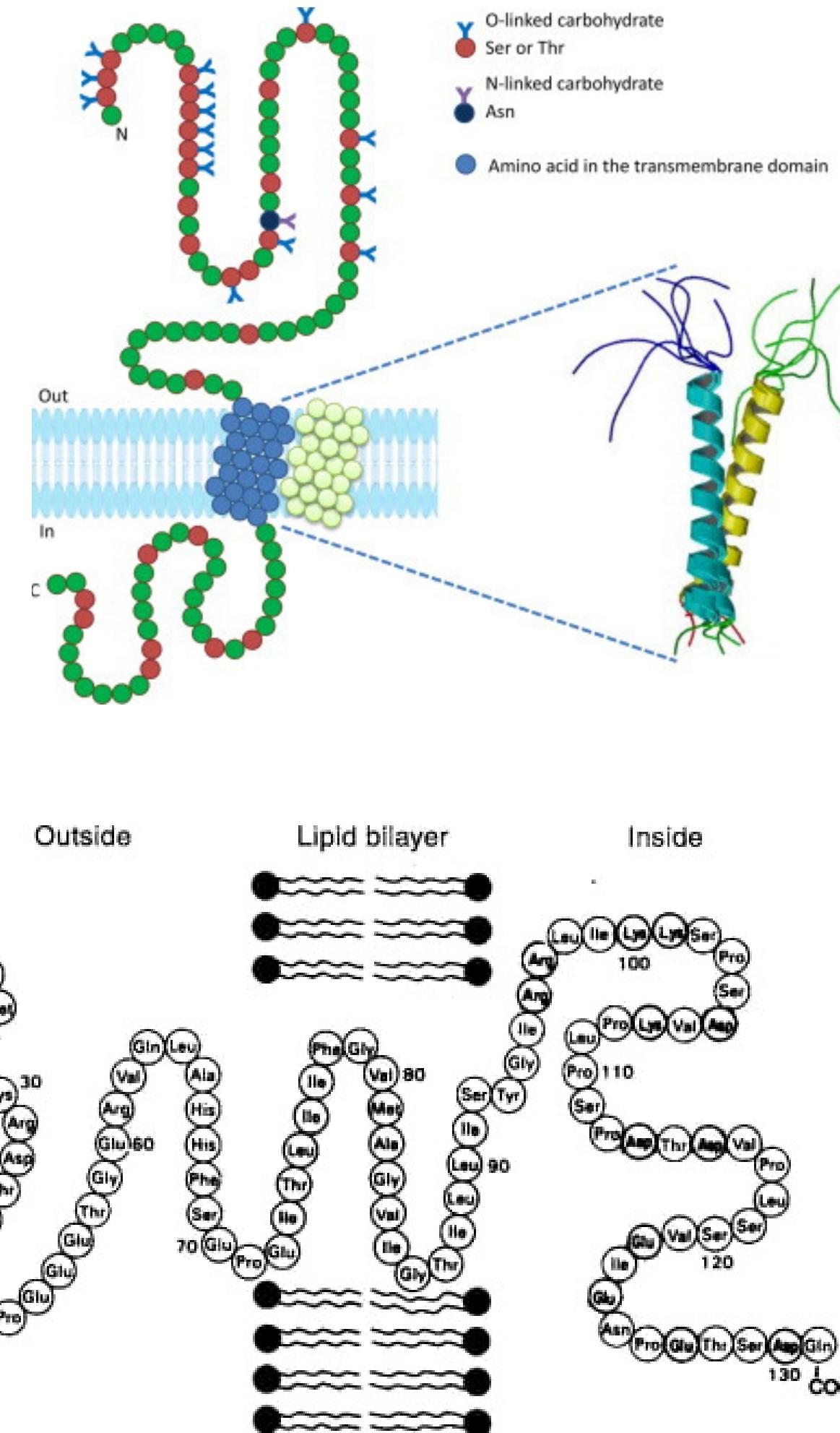
INTRODUCTION

Glycophorin A (GPA) is the major **sialoglycoprotein** on the red blood cell (RBC) membrane, present in about 0.5 million copies per cell. Encoded by the **GYPA** gene on **chromosome 4**, it constitutes nearly 2% of total RBC membrane protein mass. GPA is a **component of the ankyrin-1 complex**, a multiprotein complex involved in the **stability and shape of the erythrocyte membrane**, mediates cell-cell and **pathogen interactions**, and carries the **MN blood group antigens**. It also serves as a receptor for *Plasmodium falciparum*, the malaria parasite.

Structurally, GPA consists of three main regions:

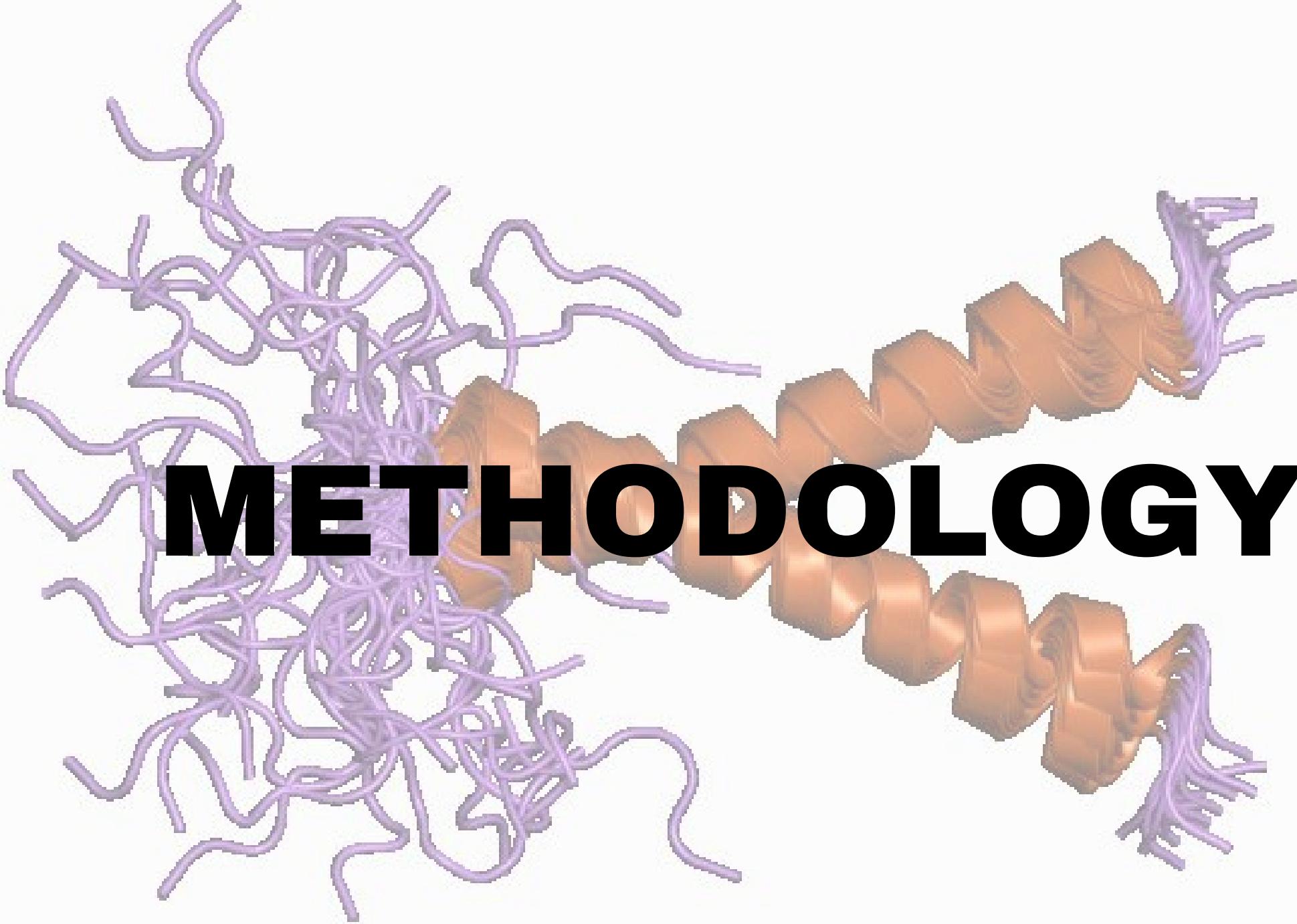
- **N-terminal extracellular segment** (residues 1–62): Heavily **O-glycosylated** on **serine** and **threonine** residues, **rich in sialic acid**, and responsible for most of the molecule's carbohydrate content.
- **Transmembrane segment** (~20–34 residues): **Hydrophobic**, spanning the lipid bilayer, and **supporting dimer formation and interactions** with other membrane proteins. Some hydrophilic residues present can possibly form salt bridges or hydrogen bonds.
- **C-terminal cytoplasmic segment** (residues 97–131): Rich in **acidic amino acids** and **proline residues**, contributing to intracellular flexibility and signaling.

The **transmembrane region** is evolutionarily conserved across mammals, whereas the N-terminal domain varies, reflecting its role in immune recognition and ligand binding.



Reasons for selecting this protein as Protein of Interest

- **Major erythrocyte membrane glycoprotein:** Highly expressed, with **- 10^6 copies per cell**, making it an ideal model for studying transmembrane protein folding, dimerization, and trafficking.
- **Blood group antigenicity:** Carries the MN and Ss blood group antigens, crucial in transfusion medicine and immunohematology.
- **Pathogen interactions:** Serves as the receptor for malarial parasite and certain viruses and bacteria, implicating it in host-pathogen recognition.
- **Heavily glycosylated extracellular domain:** Rich in O-linked glycans terminating in sialic acid, conferring a strong negative charge to the RBC surface, which prevents aggregation of erythrocytes and non-specific interactions.
- **Model for membrane protein interactions:** The transmembrane α -helix of glycophorin A is a classical example of helix-helix association via the **GxxxG motif**, making it an archetype for studying protein dimerization in lipid bilayers.
- **Clinical and evolutionary importance:** Variants of glycophorin A (such as GYPA-GYPB hybrids) confer resistance to *P. falciparum*, highlighting its role in natural selection and host adaptation.

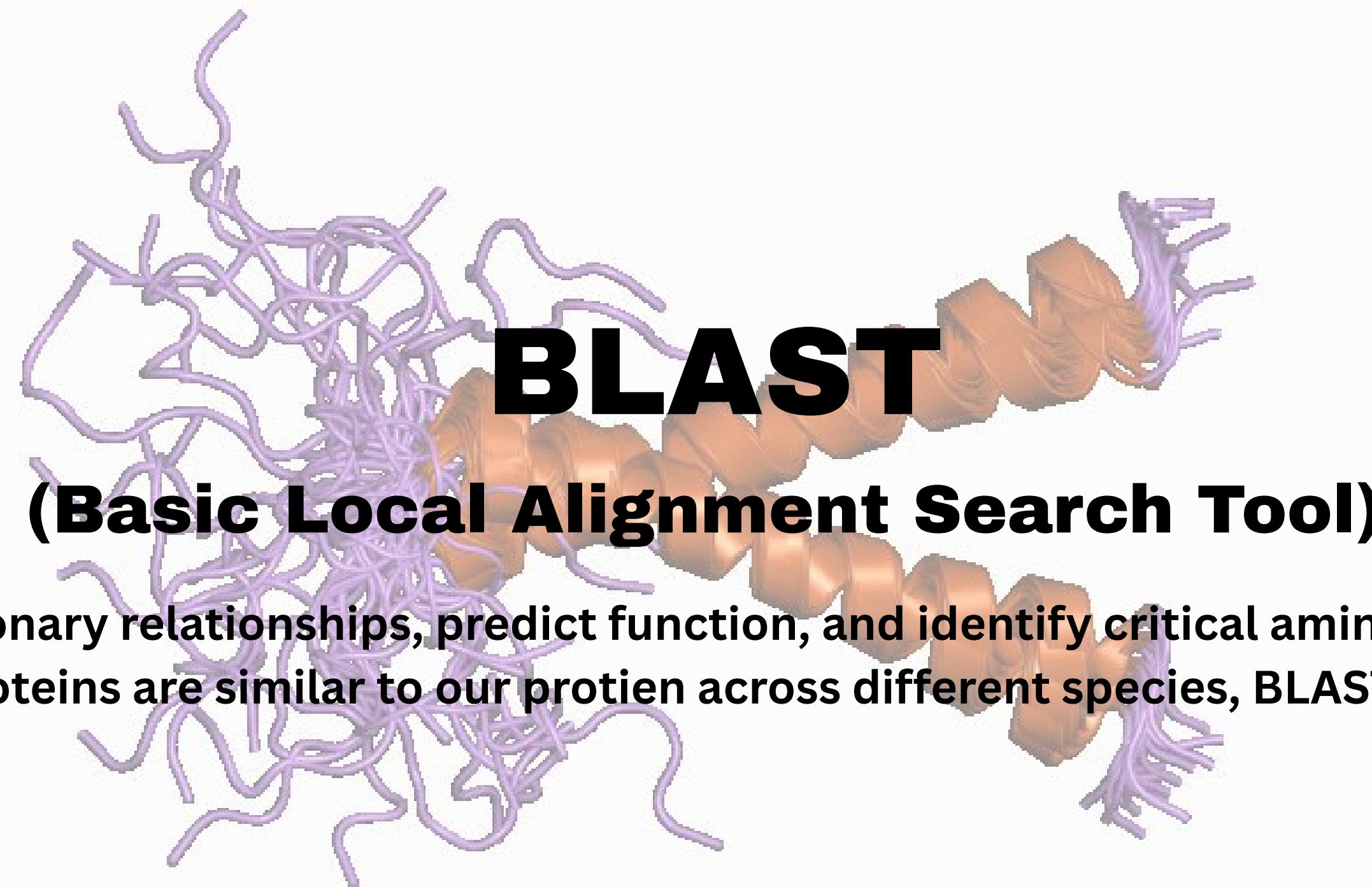


METHODOLOGY

- 1. Selection** of the protein of interest.
- 2. Identification** of the protein in the UNIPROT with its accession code.
- 3. Obtaining details** about the target Proteins on- Source Organism, Subcellular Localisation, Structure, Post Translational Model/Processing, Sequence, Family & Domains.
- 4. BLAST Search** of the entire sequence of target protein against Protein Data Bank (PDB) at NCBI-BLAST (Protein).
- 5. Multiple Sequence Alignment (MSA)** of the entire sequence of target protein with sequences obtained via BLAST Search at MultAlin and identifying conserved and non-conserved residues.
- 6. Generation of Phylogenetic Tree** using MUSCLE.
- 7. Structural Alignment in PyMOL:** between two protein structures.
- 8. Results analysed.**

Websites used for data retrieval:

- **UniProt**- Used for obtaining the protein sequence in FASTA format and other information.
<https://www.uniprot.org>
- **Pfam**- Used for obtaining information about domains of proteins.
<https://pfam.xfam.org>
- **InterPro**- InterPro provides functional analysis of proteins with respect to domain, family, homologous superfamily. The Gene Ontology (GO) terms are present as indicators of biological process, molecular function, and cellular component of a protein.
<https://www.ebi.ac.uk/interpro>
- **NCBI BLAST**- Used for finding sequences similar to target protein sequence.
<https://www.ncbi.nlm.nih.gov/BLAST>
- **MultAlin**- Used for Multiple Sequence Alignment (MSA).
<http://multalin.toulouse.inra.fr/multalin>
- **MUSCLE**- Used for generating the phylogenetic tree.
<https://www.ebi.ac.uk/Tools/msa/muscle>



BLAST

(Basic Local Alignment Search Tool)

To infer evolutionary relationships, predict function, and identify critical amino acids, and check what other proteins are similar to our protein across different species, BLAST search is done

BLAST Parameters: we used the protein BLAST (blastp) algorithm against the swissprot database,PDB database,ref sequence database.

BLAST search using PDB database

The results indicate that the potential structure and function of your query protein based on these similar, known structures.

<input checked="" type="checkbox"/>	glycophorin-A isoform X3 [Pan paniscus]	Pan panis...	187	187	100%	5e-58	68.00%	117	XP_057158153.2
<input checked="" type="checkbox"/>	PREDICTED: glycophorin-A isoform X3 [Mandrillus leucophaeus]	Mandrillus...	188	188	100%	5e-58	69.33%	149	XP_011842543.1
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Macaca thibetana thibetana]	Macaca th...	188	188	92%	6e-58	72.46%	153	XP_050647040.1
<input type="checkbox"/>	glycophorin-A isoform X3 [Trachypithecus francoisi]	Trachypith...	186	186	94%	3e-57	70.21%	153	XP_033071257.1
<input type="checkbox"/>	glycophorin-A isoform 6 [Homo sapiens]	Homo sap...	183	183	100%	1e-56	69.33%	104	NP_001425555.1
<input type="checkbox"/>	glycophorin-A isoform X1 [Rhinopithecus roxellana]	Rhinopithe...	185	185	94%	1e-56	70.21%	153	XP_030781211.1
<input checked="" type="checkbox"/>	glycophorin A precursor, partial [Gorilla gorilla]	Gorilla gor...	184	184	81%	1e-56	85.25%	123	AAB81210.1
<input type="checkbox"/>	glycophorin-A isoform X3 [Macaca fascicularis]	Macaca fa...	185	185	92%	1e-56	71.74%	153	XP_005556051.2
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Hylobates moloch]	Hylobates ...	184	184	100%	2e-56	71.33%	136	XP_032027938.1

BLAST search using refseq_protein database

The results indicate that the query sequence shows statistically significant local similarities to specific, curated and non-redundant sequences with Ref_seq

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	glycophorin-A isoform 1 precursor [Homo sapiens]	Homo sapiens	298	298	100%	8e-102	100.00%	150	NP_002090.4
<input type="checkbox"/>	glycophorin-A isoform 2 precursor [Homo sapiens]	Homo sapiens	260	260	100%	7e-87	91.33%	137	NP_001295116.1
<input type="checkbox"/>	glycophorin-A isoform 7 [Homo sapiens]	Homo sapiens	250	250	83%	3e-83	100.00%	124	NP_001425556.1
<input checked="" type="checkbox"/>	glycophorin-A precursor [Pan troglodytes]	Pan troglodytes	236	236	100%	3e-77	92.00%	150	NP_001414902.1
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Pan paniscus]	Pan paniscus	233	233	100%	4e-76	86.67%	149	XP_003815879.5
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Pongo abelii]	Pongo abelii	230	230	99%	8e-75	80.54%	167	XP_009238620.3
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Pongo pygmaeus]	Pongo pygmaeus	230	230	100%	1e-74	80.67%	160	XP_054342033.1
<input type="checkbox"/>	glycophorin-A isoform 3 precursor [Homo sapiens]	Homo sapiens	221	221	100%	6e-72	78.00%	117	NP_001295119.1
<input type="checkbox"/>	glycophorin-A isoform 4 precursor [Homo sapiens]	Homo sapiens	221	221	89%	1e-71	89.55%	145	NP_001424975.1
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	218	218	100%	3e-70	84.00%	149	XP_055240733.2
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Papio anubis]	Papio anubis	214	214	100%	2e-68	75.33%	162	XP_021794517.1
<input type="checkbox"/>	PREDICTED: glycophorin-A isoform X1 [Mandrillus leucophaeus]	Mandrillus leucophaeus	213	213	100%	5e-68	75.33%	162	XP_011842541.1
<input type="checkbox"/>	glycophorin-A isoform X3 [Pongo abelii]	Pongo abelii	212	212	92%	8e-68	79.71%	146	XP_054410489.1
<input type="checkbox"/>	glycophorin-A isoform X2 [Pongo abelii]	Pongo abelii	212	212	92%	1e-67	79.71%	154	XP_054410488.1
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Macaca thibetana thibetana]	Macaca thibetana thibetana	212	212	100%	1e-67	74.67%	162	XP_050647039.1
<input type="checkbox"/>	PREDICTED: glycophorin-A isoform X1 [Cercopithecus atys]	Cercopithecus atys	211	211	100%	3e-67	74.67%	162	XP_011904573.1
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Symphalangus syndactylus]	Symphalangus syndactylus	209	209	100%	7e-67	77.33%	149	XP_055130020.1
<input type="checkbox"/>	glycophorin-A isoform X1 [Pan troglodytes]	Pan troglodytes	207	207	83%	2e-66	91.94%	124	XP_054539847.1
<input type="checkbox"/>	glycophorin-A isoform X1 [Macaca fascicularis]	Macaca fascicularis	209	209	100%	3e-66	74.00%	162	XP_005556050.2
<input checked="" type="checkbox"/>	glycophorin-A isoform X1 [Macaca mulatta]	Macaca mulatta	208	208	100%	5e-66	73.33%	162	XP_014994775.2

BLAST search using non-redundant protein database

The blast results explains a query protein's likely function, evolutionary relationships, and potential source organism by finding and ranking similar sequences in the database.

<input checked="" type="checkbox"/>	glycophorin-A isoform X3 [Theropithecus gelada]	Theropithe...	189	189	100%	2e-58	69.33%	149	XP_025241
<input type="checkbox"/>	PREDICTED: glycophorin-A isoform X2 [Mandrillus leucophaeus]	Mandrillus...	189	189	95%	3e-58	71.83%	160	XP_011842
<input checked="" type="checkbox"/>	glycophorin-A isoform X3 [Pan paniscus]	Pan panis...	187	187	100%	5e-58	68.00%	117	XP_057158
<input checked="" type="checkbox"/>	PREDICTED: glycophorin-A isoform X3 [Mandrillus leucophaeus]	Mandrillus...	188	188	100%	5e-58	69.33%	149	XP_011842
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Macaca thibetana thibetana]	Macaca th...	188	188	92%	6e-58	72.46%	153	XP_050647
<input type="checkbox"/>	glycophorin-A isoform X3 [Trachypithecus francoisi]	Trachypith...	186	186	94%	3e-57	70.21%	153	XP_033071
<input type="checkbox"/>	glycophorin-A isoform 6 [Homo sapiens]	Homo sap...	183	183	100%	1e-56	69.33%	104	NP_001425
<input type="checkbox"/>	glycophorin-A isoform X1 [Rhinopithecus roxellana]	Rhinopithe...	185	185	94%	1e-56	70.21%	153	XP_030781
<input checked="" type="checkbox"/>	glycophorin A precursor, partial [Gorilla gorilla]	Gorilla gor...	184	184	81%	1e-56	85.25%	123	AAB81210
<input type="checkbox"/>	glycophorin-A isoform X3 [Macaca fascicularis]	Macaca fa...	185	185	92%	1e-56	71.74%	153	XP_005556
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Hylobates moloch]	Hylobates ...	184	184	100%	2e-56	71.33%	136	XP_032027
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Macaca fascicularis]	Macaca fa...	184	184	92%	2e-56	71.74%	160	XP_045248
<input checked="" type="checkbox"/>	glycophorin-A isoform X2 [Macaca mulatta]	Macaca m...	184	184	92%	3e-56	71.01%	153	XP_028704

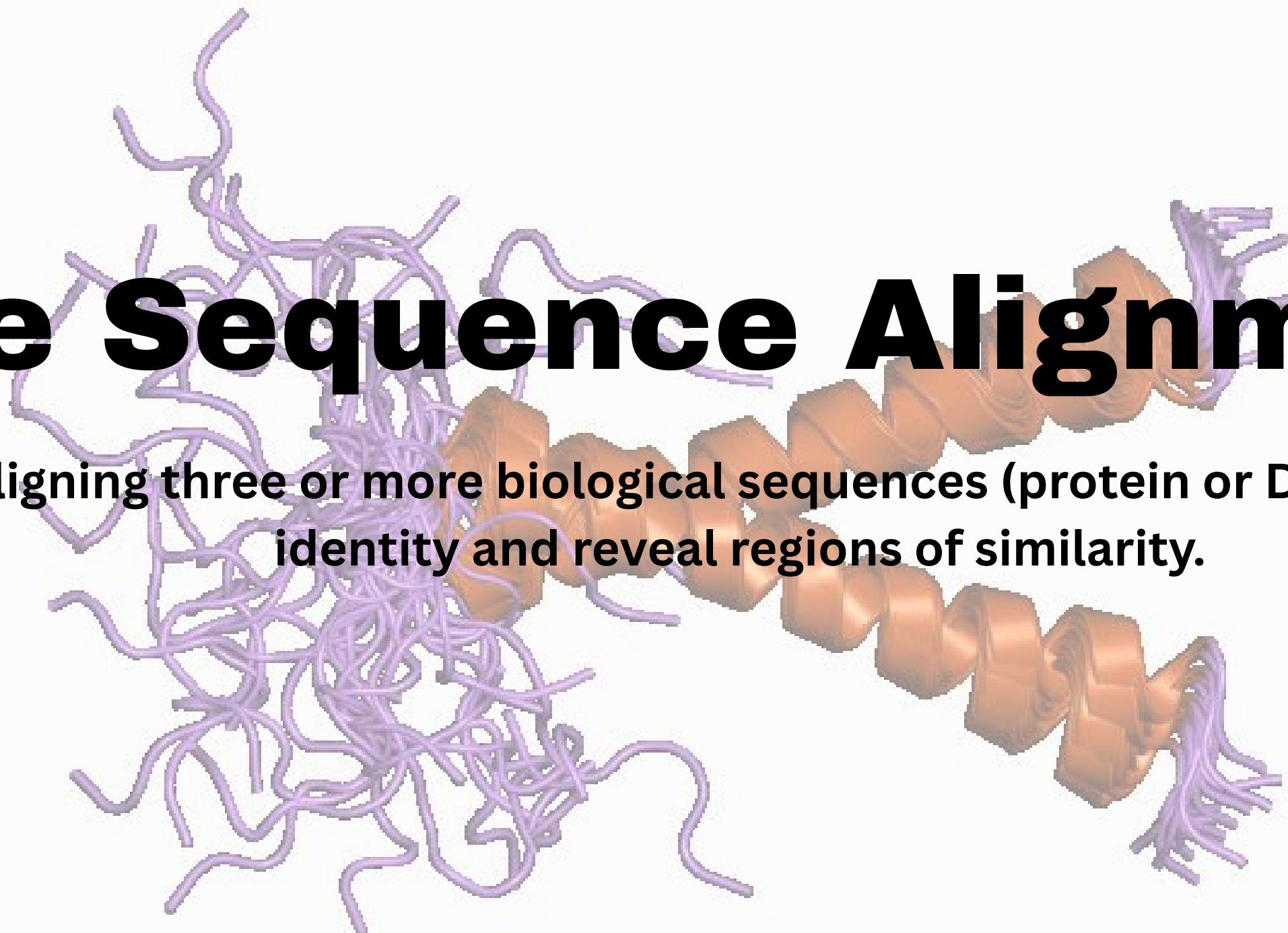
BLAST search using Swissprot database

The results can give us information about the detailed function of our query protein.

Sequences producing significant alignments											Download	Select columns	Show 100	?
		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession				
<input type="checkbox"/> select all 8 sequences selected														
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-A; AltName: Full=MN sialoglycoprotein; AltName: Full=PAS-2; AltName: Full=Sialoglyco...		Homo sapiens	298	298	100%	9e-105	100.00%	150	P02724.3					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-A; AltName: CD_antigen=CD235a; Flags: Precursor [Pan troglodytes]		Pan troglodytes	241	241	100%	3e-82	89.33%	149	Q28913.1					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-B; AltName: CD_antigen=CD235b; Flags: Precursor [Pan troglodytes]		Pan troglodytes	162	162	79%	1e-51	81.97%	123	Q28914.1					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-A; AltName: Full=Glycophorin-MK; AltName: CD_antigen=CD235a [Macaca fuscata fus...		Macaca fuscata f...	154	154	87%	7e-48	64.89%	144	P14221.1					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-B; AltName: Full=PAS-3; AltName: Full=SS-active sialoglycoprotein; AltName: Full=Sial...		Homo sapiens	102	102	79%	4e-28	62.30%	91	P06028.3					
<input type="checkbox"/> RecName: Full=Glycophorin-A; AltName: Full=Glycophorin-HA; AltName: CD_antigen=CD235a [Equus caballus]		Equus caballus	73.6	73.6	49%	2e-16	53.25%	120	P02726.1					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-E; Flags: Precursor [Homo sapiens]		Homo sapiens	64.3	64.3	77%	2e-13	48.70%	78	P15421.2					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-A; AltName: CD_antigen=CD235a [Mus musculus]		Mus musculus	62.4	62.4	45%	1e-11	51.47%	168	P14220.2					
<input type="checkbox"/> RecName: Full=Glycophorin-A; AltName: CD_antigen=CD235a [Sus scrofa]		Sus scrofa	57.8	57.8	48%	3e-10	44.83%	133	P02725.1					
<input checked="" type="checkbox"/> RecName: Full=Glycophorin-E; Flags: Precursor [Pan troglodytes]		Pan troglodytes	55.5	55.5	79%	7e-10	52.10%	78	Q28915.1					

Multiple Sequence Alignment (MSA)

The process of aligning three or more biological sequences (protein or DNA) to maximize residue identity and reveal regions of similarity.



MSA using Multalin

1. [for results obtained from non redundant database blast search for a selected number of closely related species of human]

	1	10	20	30	40	50	60
NP_002090.4:1-150	MYGKIIIFVLLI	S	EIVSISALSTTEVAMHTSTSSSVT	K	SYISSQT	N	DTHKRDTYAAATPRAH
Q28913.1:1-149	MYGKIIIFVLLI	S	AIVSISASSTTEVAMHTSTSS.VT	K	SYISSQT	S	DKHKWDTYPATPRAH
XP_021794517.1:1-150	MYGKIIIFVLLI	S	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTYPTTFSAH
XP_011904573.1:1-150	MYGKIIIFVLLI	S	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	P	NDKHTSDTHPTTFSAH
XP_050647040.1:4-141	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTHPTTFSAH
XP_045248845.1:11-14	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTHPTTFSAH
XP_028704616.1:4-141	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTHPTTFSAH
XP_070949081.1:4-141	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTHPTTASAH
XP_023081750.2:1-153	MYGKIIIFVLLI	S	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	P	NDKHTSDTHPTTFSAH
XP_023081751.2:13-15	C	EIVSISASSTTVPAMHTSTSSLVF	G	SYVSSQ	P	NDKYKWDIHPTPSAH
XP_033071259.1:4-144	EIVSISASSTPVPEMHTSTSSLVF	E	SYVSSQ	P	NDKHKWDIHPTPSAH
XP_030781227.1:4-141	EIVSISASSTTVPEMHTSTSSLVF	G	SYVSSR	P	NDKHKWDIHPHTPSAH
KAL4843006.1:1-138	L	EIVSISASSTTVPEMHTSTSSLVF	G	SYVSSQ	P	NDKHKWDIHPHTPSAH
XP_025241177.1:1-137	MYGKIIIFVLLI	S	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	S	NDKHTSDTYPTTFSAH
XP_011842543.1:1-137	MYGKIIIFVLLI	S	EIVRISASSTTVPATHTSTSSLVF	E	SYVSSQ	P	NDKHTSDTHPTTFSAH
XP_055130020.1:1-149	MYEKIKFVLLI	L	EIASISAPSTTGLVMHTSISSVT	K	SYTSSQT	T	NDKNKWYTYPAR.SVN
XP_032027938.1:1-136	MYEKIKFVLLI	L	EIASISALSTTGLVMHASISSVT	K	SYTSSQT	T	NDKNKWYTYPAR.SVN
XP_055130021.1:1-136	MYEKIKFVLLI	L	EIASISAPSTTGLVMHTSISSVT	K	SYTSSQT	T	NDKNKWYTYPAR.SVN
XP_012363399.1:1-136	MYEKIKFVLLI	F	EIASISSTTGLVMHTSISSVT	K	SYTSSQT	T	NDKNKWYTYPAR.GVN
PNJ83724.1:1-137	MYEKIKFVLLI	S	EIVSIPASNTTGEVMHTSISSVT	K	SYITPQT	T	NDKHKQDTYATPFSAH
AAB81210.1:1-122	L	EIVSISASSTTEVAMHTSTSSVT	K	SYISSET	S	NDKHKRDTYAAPFSAH
XP_057158153.2:1-117	MYGKIIIFVLLI	S	DKQKWDTYATPRAH		

	70	80	90	100	110	120
NP_002090.4:1-150	EVSEISVRTVYPPE	E	TGERVQLAHHFSEPE	E	LIIIFGVMAVGIVGTLIL	SYCIRRRIKK
Q28913.1:1-149	EVSEIYVTTVYPPE	E	ENGEGVQLVHRFSEPE	E	LIIIFGVMAVGIVGTLIL	IYYSIRRRIKK
XP_021794517.1:1-150	EVSGFSGRTHYPPE	E	EDNRERQLVHFRSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_011904573.1:1-150	EVSEFSGRTHYPPE	E	EDNRERQLVHFRSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_050647040.1:4-141	EVSEFSGRTHYPPE	E	EDNGERVQLVHEFSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_045248845.1:11-14	EVSEFSGRTHYPPE	E	EDNRERVQLVHEFSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_028704616.1:4-141	EVSEFSGRTHYPPE	E	EDNRERVQLVHEFSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_070949081.1:4-141	EVSEFSGRTHYPPE	E	EDNGERVQLVHEFSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_023081750.2:1-153	EVLEFPGRNHYPP	E	ENRKREQLVHEFSELV	E	LIIIFGVMAVVIGTILFISYC	IRRDRKK
XP_023081751.2:13-15	EVLEFPGRNHYPP	E	ENRKREQLVHEFSELV	E	LIIIFGVMAVVIGTILFISYC	IRRDRKK
XP_033071259.1:4-144	EVSEFSGRNYHPP	E	ENRKREQLVHEFSELV	E	LIIILGVMAVGIVGTLIFISYC	IRRDRKK
XP_030781227.1:4-141	EVSEIYVTPHYTE	E	ENREREQLVHEFSELV	E	LIIILGVMAVGIVGTLIFISYC	IRRDRKK
KAL4843006.1:1-138	EVSEFSGRNYHPIE	E	ENREREQLVHEFSELV	E	LIIIFGVMAVGIVGTLIFISYC	IRRDRKK
XP_025241177.1:1-137	EVSGFSGRTHYPPE	E	EDN	MALIIIFGVMAVGIVGTLIFISYC
XP_011842543.1:1-137	EVSEFSGRTRYPP	E	EDN	MALIIIFGVMAVGIVGTLIFISYC
XP_055130020.1:1-149	EVSEISVTTVYPPE	E	EDGENRQLVHRFSEPE	E	LIIIFGVMAVGIVGTLISIYC	IRLDRKK
XP_032027938.1:1-136	EVSEISVTTVYPPE	E	EDGENRQLVHRFSEPE	E	VITLIIFGVMAVGIVGTLISIYC	IRLDRKK
XP_055130021.1:1-136	EVSEISVTTVYPPE	E	EDGENRQLVHRFSEPE	E	VITLIIFGVMAVGIVGTLISIYC	IRLDRKK
XP_012363399.1:1-136	EVSEISVTTVHPPE	E	EDGENRQLVHRFSEPE	E	VITLIIFGVMAVGIVGTLISIYC	IRLDRKK
PNJ83724.1:1-137	EVSEISVTTIHSPE	E	ENR	VITLIFGVMAVGIVGTLISIYC
AAB81210.1:1-122	EVSEISVAVYPPPE	E	NGERVQLVHRFSEAE	E	LIIIFGVMAGIIGTILFISY	IRRDRKK
XP_057158153.2:1-117	EVSEIYVTTVYPPE	E	ENGERQLVHRFQP	E	LIIIFGVMAVGIVGTLIL	IYYSICRDRKK

	130	140	150	
NP_002090.4:1-150	SPSDVKEP	E	SPDTDVPPLSSVEIENP...	EETSDQ
Q28913.1:1-149	SPSDVKEP	E	SPDTDVPPLSSVEIENP...	EETSDQ
XP_021794517.1:1-150	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_011904573.1:1-150	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_050647040.1:4-141	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_045248845.1:11-14	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_028704616.1:4-141	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_070949081.1:4-141	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_023081750.2:1-153	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_023081751.2:13-15	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_033071259.1:4-144	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
KAL4843006.1:1-138	SPSDVKEP	E	SPPDAEVPLSSVEIENP...	EETDQ
XP_025241177.1:1-137	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_011842543.1:1-137	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_055130020.1:1-149	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_032027938.1:1-136	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_055130021.1:1-136	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_012363399.1:1-136	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
PNJ83724.1:1-137	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
AAB81210.1:1-122	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ
XP_057158153.2:1-117	SPSDVKEP	E	SPPDTEVPPLSSVEIENP...	EETDQ

MSA using Multalin

2. [for results obtained from non redundant database blast search of variants of human Glycophorin A]

	1	10	20	30	40	50	60
UZP80292.1:1-150	MYGKIIIFVLLLSAIVSISALSTTEVEMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
UZP80290.1:1-150	MYGKIIIFVLLLSAIVSISASSTTGVAHMHSSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
AAX53135.1:1-150	MYGKIIIFVLLLSAIVSISASSTTGVAHMHSSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
NP_001414902.1:1-150	MYGKIIIFVLLSEIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
BCH47620.1:1-149	MYGKIIIFVLLSEIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
NP_001295116.1:1-137	MYGKIIIFVLLSEIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
EAX05060.1:1-137	MYGKIIIFVLLLSAIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
KAI4027145.1:1-137	MYGKIIIFVLLLSAIVSISASSTTGVAHMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
NP_001424975.1:1-134	MYGKIIIFVLLSEIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
KAI4027138.1:1-134	MYGKIIIFVLLLSAIVSISASSTTGVAHMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
ADU25344.1:1-134	MYGKIIIFVLLLSAIVSISASSTTGVAHMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
AAB34057.1:1-122	MYGKIIIFVLLLSAIVSISALSTTEVAMHTSTSSSVT	KSYISSQNTNDTHKRDTYAAATPRAH					
NP_001295119.1:1-117
NP_001425555.1:1-104
KAI2536045.1:1-118	MYGKIIIFVLLSEIVSISALSTTEVAMHTSTSSSVT	KSYISSQTN.....					
AAC50058.1:1-118	MYGKIIIFVLLLSAIVSISALSTTEVAMHTSTSSSVT	KSYISSQTN.....					
KAI2536044.1:1-105	MYGKIIIFVLLLSAIVSISALSTTEVAMHTSTSSSVT	KSYISSQTN.....					
EAX05059.1:1-105	MYGKIIIFVLLLSAIVSISALSTTEVAMHTSTSSSVT	KSYISSQTN.....					
consensus>50	mygkiiifvlllsai...salsttevamhtstsssvt	Ksyissqntndthkrdt...aatprah					

	70	80	90	100	110
UZP80292.1:1-150	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
UZP80290.1:1-150	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
AAX53135.1:1-150	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
NP_001414902.1:1-150	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
BCH47620.1:1-149	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
NP_001295116.1:1-137	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
EAX05060.1:1-137	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
KAI4027145.1:1-137	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
NP_001424975.1:1-134	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
KAI4027138.1:1-134	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
ADU25344.1:1-134	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
AAB34057.1:1-122	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
NP_001295119.1:1-117	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
NP_001425555.1:1-104	EVSEISVRTVYPPEEETGERVQLAHHFSEPE	ITLIIIFGV...MAGVIGTILLISY...GIRRL			
KAI2536045.1:1-118
AAC50058.1:1-118
KAI2536044.1:1-105
EAX05059.1:1-105
consensus>50	evseisvrtvyppeeetgervqlahhfsepe...	ItLIIIfgv...magvigtillisy...girrl			

	120	130	140	150
UZP80292.1:1-150	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
UZP80290.1:1-150	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
AAX53135.1:1-150	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
NP_001414902.1:1-150	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
BCH47620.1:1-149	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
NP_001295116.1:1-137	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
EAX05060.1:1-137	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
KAI4027145.1:1-137	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
NP_001424975.1:1-134	IKRQVINENLFTKP...NVE			
KAI4027138.1:1-134	IKRQVINENLFTKP...NVE			
ADU25344.1:1-134	IKRQVINENLFTKP...NVE			
AAB34057.1:1-122	IKRQVINENLFTKP...NVE			
NP_001295119.1:1-117	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
NP_001425555.1:1-104	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
KAI2536045.1:1-118	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
AAC50058.1:1-118	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
KAI2536044.1:1-105	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
EAX05059.1:1-105	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			
consensus>50	IKKSPSDVKPLPSPDTDVPLSSVEIENPETSQ			

MSA using Multalin

[for results obtained from refseq_protien database blast search]

	1	10	20	30	40	50	60																																																							
NP_002090.4:1-150	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	S	I	S	A	L	S	T	T	E	V	A	M	H	T	S	T	S	S	V	T	K	S	Y	I	S	S	Q	T	N	D	T	H	K	R	D	T	Y	A	A	T	P	R	A					
NP_001414902.1:1-150	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	S	I	S	A	S	S	T	T	E	V	A	M	H	T	S	T	S	S	V	T	K	S	Y	I	S	S	E	T	N	D	K	H	K	R	D	T	Y	A	A	T	P	R	A					
XP_003815879.5:1-149	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	S	I	S	A	S	S	T	T	E	V	A	M	H	T	S	T	S	S	.	V	T	K	S	Y	I	S	S	E	T	S	D	K	Q	K	W	D	T	Y	P	A	T	P	R	A				
XP_055240733.2:1-149	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	S	I	S	A	S	S	T	T	E	V	A	V	H	T	S	T	S	S	V	T	K	S	Y	I	S	S	Q	T	N	D	K	H	K	Q	D	T	Y	P	A	H	.	V	N					
XP_009238620.3:19-16	.	Y	E	K	I	I	F	V	L	L	S	E	I	V	S	I	P	A	S	N	T	T	E	V	A	M	H	T	S	I	S	S	V	T	K	S	Y	I	T	P	Q	T	N	D	K	H	K	Q	D	T	Y	P	A	H	.	V	N					
XP_055130020.1:1-149	M	Y	E	K	I	K	F	V	L	L	L	E	I	A	S	I	S	A	P	S	T	T	E	V	A	M	H	T	S	I	S	S	V	T	K	S	Y	T	T	S	Q	T	N	D	K	N	K	W	Y	T	Y	P	A	R	S	.	V	N				
XP_032027936.1:1-151	M	Y	E	K	I	K	F	V	L	L	L	E	I	A	S	I	S	A	L	S	T	T	E	V	A	M	H	T	S	I	S	S	V	T	K	S	Y	T	T	S	Q	T	N	D	K	N	K	W	Y	T	Y	P	A	R	S	.	V	N				
XP_021794517.1:1-150	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	R	I	S	A	S	S	T	T	E	V	A	T	H	T	S	T	S	S	L	V	P	E	S	Y	V	S	S	Q	S	N	D	K	H	T	S	D	T	Y	P	T	P	S	A	H	.	V	N	
XP_050647039.1:1-150	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	R	I	S	A	S	S	T	T	E	V	A	T	H	T	S	T	S	S	L	G	P	E	S	Y	V	S	S	Q	S	N	D	K	H	T	S	D	T	H	P	T	T	P	S	A	H	.	V	N
XP_014994775.2:1-150	M	Y	G	K	I	I	F	V	L	L	S	E	I	V	R	I	S	A	S	S	T	T	E	V	A	T	H	T	S	T	S	S	L	G	P	E	S	Y	V	S	S	Q	S	N	D	K	H	T	S	D	T	H	P	T	T	P	S	A	H	.	V	N
consensus>50	m	Y	g	K	I	I	F	V	L	L	S	e	I	v	s	I	s	a	s	s	t	t	.	a	m	h	t	s	t	s	s	v	t	k	s	y	i	s	s	q	t	n	d	t	d	h	k	.	d	t	y	p	a	p	.	a	h	.	v	n		

	70	80	90	100	110																																																								
NP_002090.4:1-150	E	V	S	E	I	S	V	R	T	V	Y	P	P	E	E	E	T	G	E	R	V	Q	.	.	L	A	H	H	F	S	E	P	E	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	G	I	R	R	L	I	
NP_001414902.1:1-150	E	V	S	E	I	S	V	T	F	V	Y	P	P	E	V	Y	N	G	E	R	V	Q	.	.	L	V	H	R	F	S	E	P	E	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	S	I	R	R	L	I	
XP_003815879.5:1-149	E	V	S	E	I	I	Y	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	P	E	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	S	I	C	R	L	I
XP_055240733.2:1-149	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	P	R	F	S	G	P	E	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	S	I	R	R	L	I
XP_009238620.3:19-16	E	V	S	E	I	S	V	T	I	H	S	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
XP_055130020.1:1-149	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
XP_032027936.1:1-151	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
XP_021794517.1:1-150	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
XP_050647039.1:1-150	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
XP_014994775.2:1-150	E	V	S	E	I	S	V	T	T	V	Y	P	P	E	E	E	N	E	G	E	R	Q	Q	.	.	L	V	H	R	F	S	E	P	V	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I
consensus>50	E	V	S	e	i	s	v	t	t	v	y	p	p	e	e	e	E	E	E	R	Q	Q	.	.	L	v	h	r	f	s	e	p	v	i	t	l	I	I	F	G	V	M	A	G	V	I	G	T	I	L	I	I	S	Y	C	I	R	R	L	I	

	120	130	140	150																												
NP_002090.4:1-150	K	K	S	P	S	D	V	K	P	L	P	S	P	D	T	D	V	P	L	S	S	V	E	I	E	N	P	E	T	S	D	Q
NP_001414902.1																																

MSA using Multalin

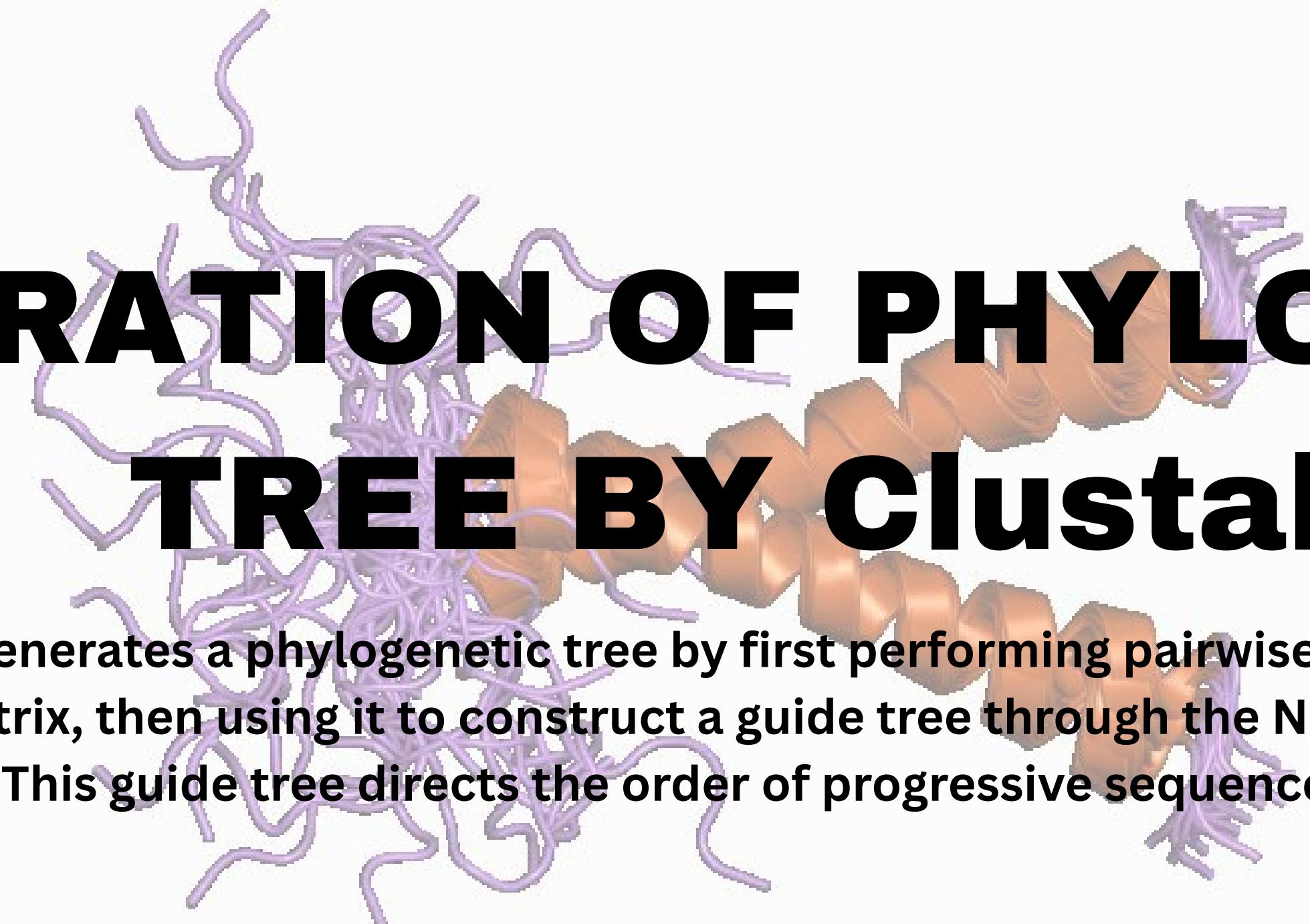
[for results obtained from Swissprot database blast search]

	1	10	20	30	40	50	60
P02724.3:1-150	MYGKIIIFVLLS	EIVSISALSTTEVAMHTSTSSSVT	KSYISSQTNDTHKR	DTYAATPRAH			
Q28913.1:1-149	MYGKIIIFVLLS	AIIVSISASSTTEVAMHTSTSS.	VTKSYISSETSDHKH	KWDTYPATPRAH			
Q28914.1:1-122	MYGKIIIFVLLS	EIVSISASSTTEVAMHTSTSSSV	TKSYSQ	ISSQTNDHKHG	DTYPATLG	AH	
P14221.1:2-130	STTVPATHTSSSSLGPEQYV	SSSQ	SNDKHTSDSHPTPTS	AH		
P06028.3:1-90	MYGKIIIFVLLS	EIVSISALSTTEVAMHTSTSSSV	TKSYSQ	ISSQTNDHKHG	DTYPATLG	AH	
Q28915.1:1-77	MYGKIIIFVLLS	AIIVSISASSTTGVAMHTSTSSSV	TKSYSQ	ISSQIN....	
P15421.2:1-78	MYGKIIIFVLLS	SGIVSISASSTTGVAMHTSTSSSV	TKSYSQ	ISSQTNDHKHG	DTYPATLG	AH	
P14220.2:100-167	
consensus>50	mygkiiifvlls.	.ivsisassttevamhtstsssvtksyissqtnd.	h..d.....	ah			

	70	80	90	100	110
P02724.3:1-150	EV	S.EISVR	TVYPPEEE	TGERVQLAH	HFSPEE
Q28913.1:1-149	EV	S.EIYVT	TVYPPEEE	ENGEGVQLVH	RFSPEE
Q28914.1:1-122	EV	S.EISVT	TVYPPEED	NGEWVQPVHP	FSRPAPV
P14221.1:2-130	EV	TTEFSGR	THYPPEED	..DRVQLVHE	FSE
P06028.3:1-90	GETGQLVH	RF
Q28915.1:1-77	TVPA	PV
P15421.2:1-78	G	VI	L
P14220.2:100-167	ALVEKIL	IILCPMAGVI
consensus>50	ev..e....t.	yppeee.ge.vqlvh.fsep...	viil!ilgvmag!!gtilli	sy	cisr

	120	130	140	150
P02724.3:1-150	LIKKS	PSDV	KPLPSPD	TDVPLSSVE
Q28913.1:1-149	LIKKS	PSDV	KPLPSPD	TDVPLSSVE
Q28914.1:1-122	LIK
P14221.1:2-130	LIKKS	ESDV	QPLPPPDAE	VPLSSVE
P06028.3:1-90	LIK
Q28915.1:1-77	LIK
P15421.2:1-78
P14220.2:100-167	MTKKS	SV	DIQSPEGGD	NSVPLSSIE
consensus>50	likks..dv.....d..	vplssve.....e...q		

GENERATION OF PHYLOGENETIC TREE BY ClustalW



ClustalW generates a phylogenetic tree by first performing pairwise alignments to create a distance matrix, then using it to construct a guide tree through the Neighbor-Joining method.

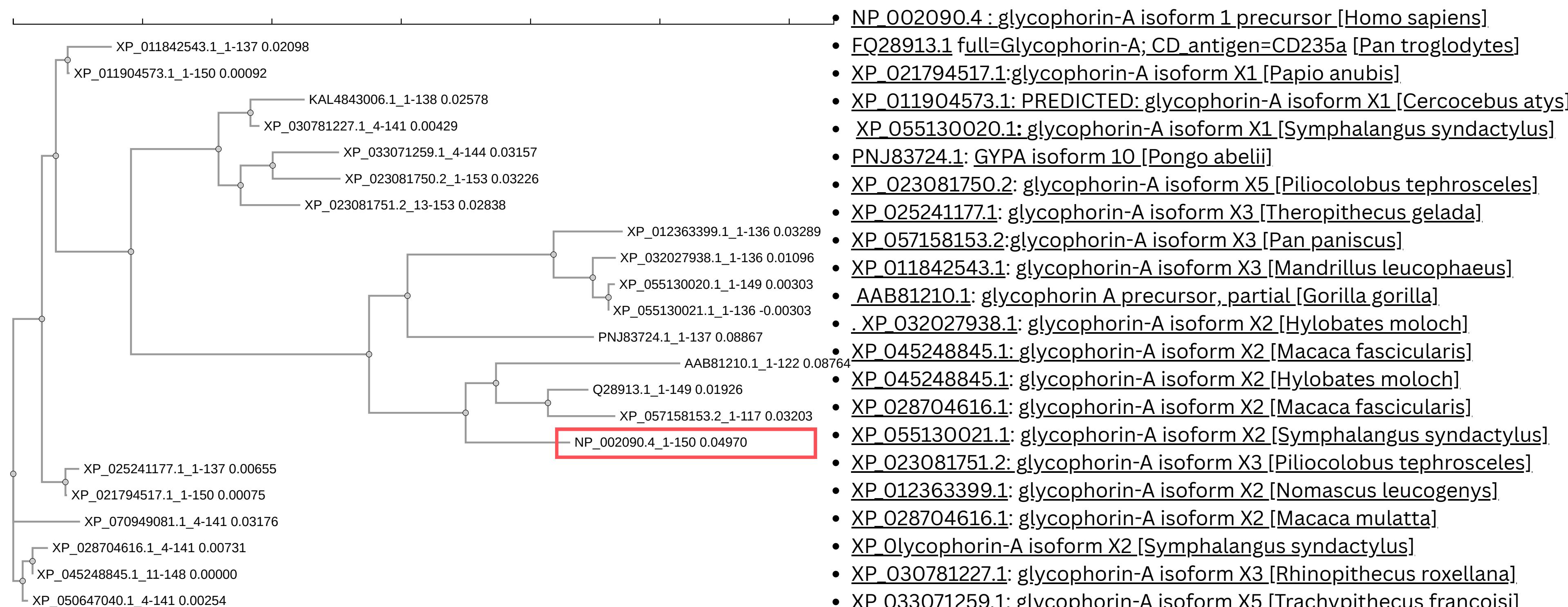
This guide tree directs the order of progressive sequence alignment.

Phylogenetic Tree of Human and related species:

For this purpose, a phylogenetic tree is generated using MUSCLE from the selected 22 sequences from non-redundant blast search.

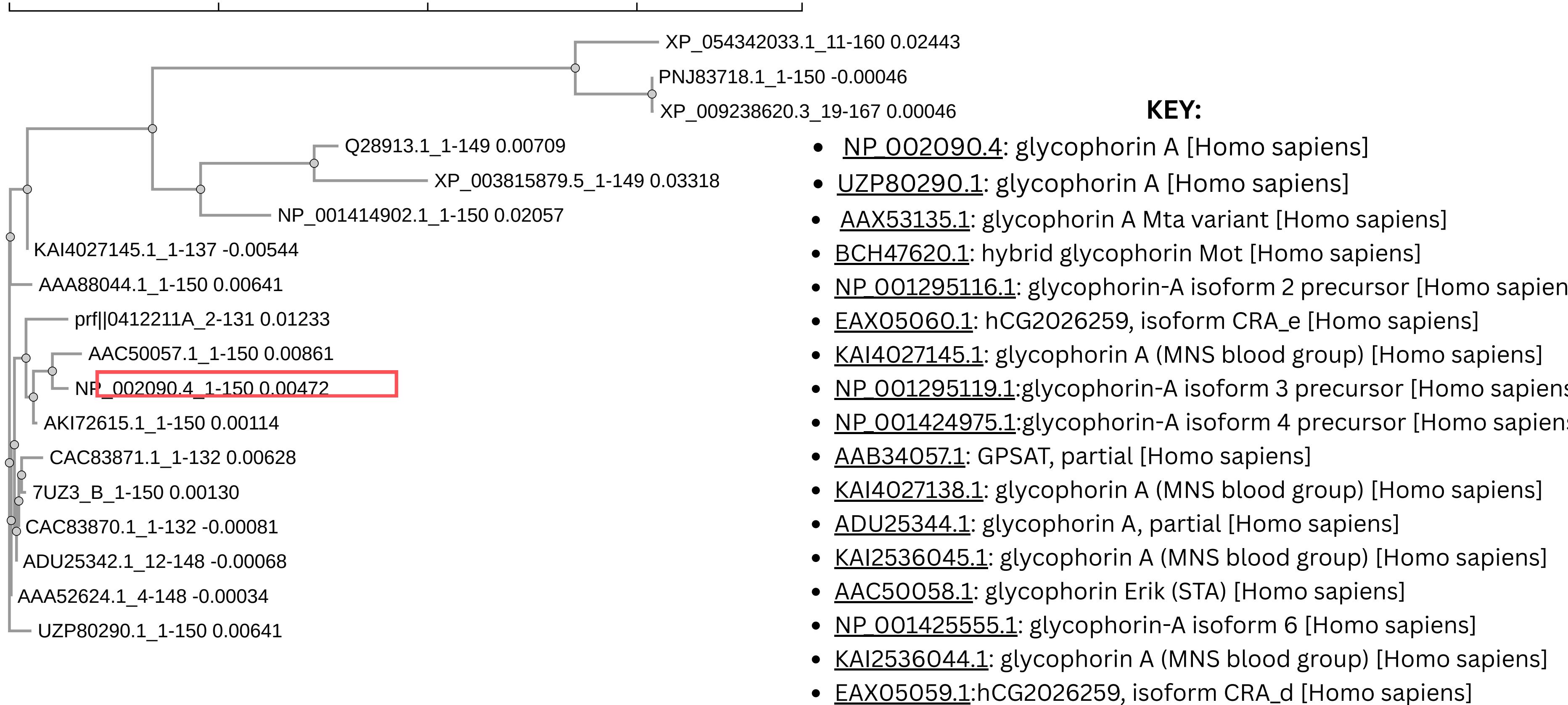
1. Phylogenetic tree of human with 21 related species found using BLAST search

KEY:



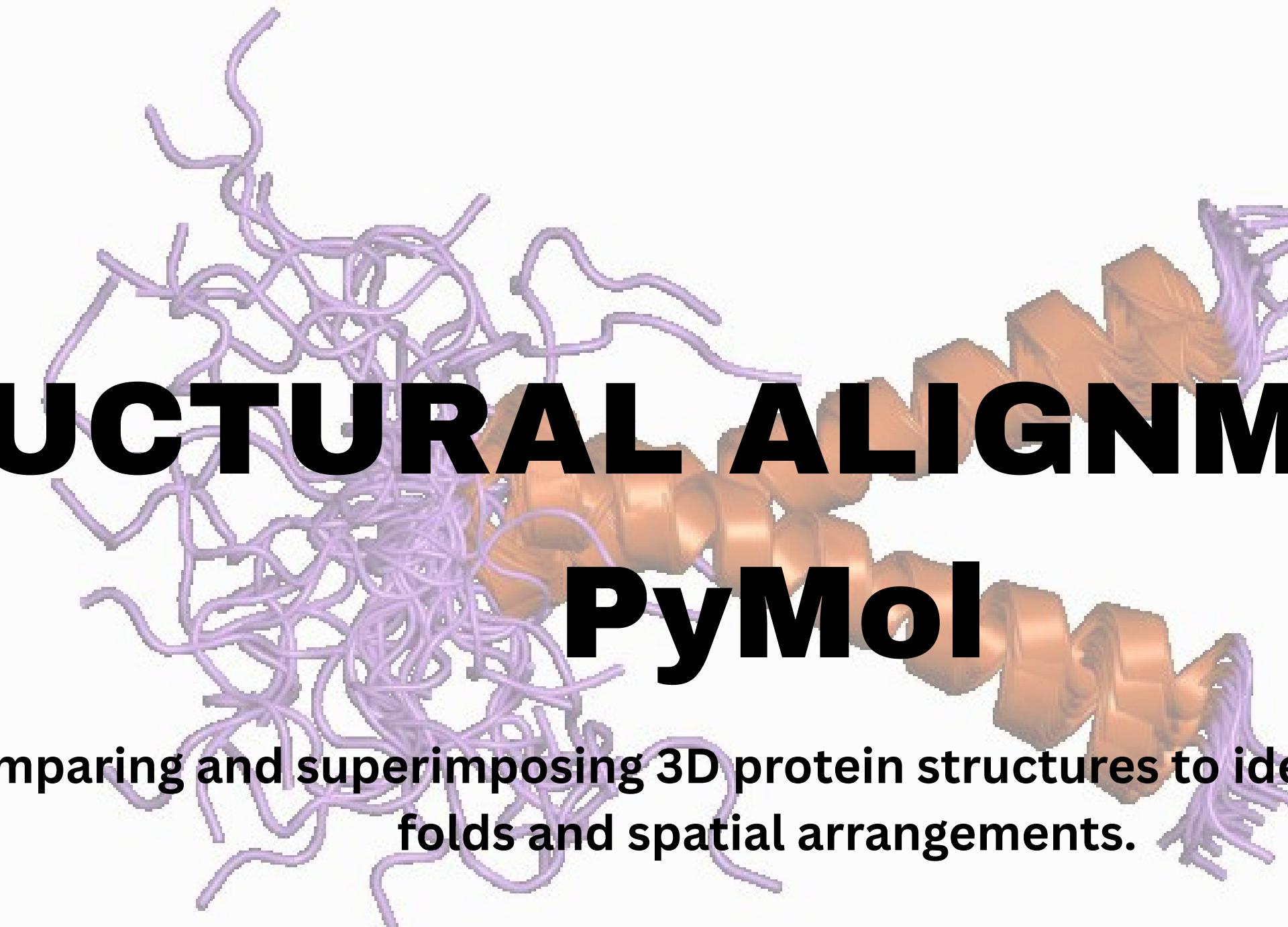
Phylogenetic Tree of Glycophorin a of human

2. Phylogenetic tree of 18 non redundant sequences of human found using BLAST search



EVOLUTIONARY SIGNIFICANCE OF THE PHYLOGENETIC TREE

- The evolutionary significance of the phylogenetic tree is concluded by **analyzing the relationships and distances between the sequences**, which represent different glycophorin A proteins. There two varieties of phylogenetic trees that has been created here. One shows the evolutionary relationships of our query sequence across different genus. The other tree gives the evolutionary relationships within the same genus but different species.
- Sequences that are **more closely related** are **grouped together** on the **same branches** of the tree.
- The **length** of the branches represents the **evolutionary distance between the sequences**. Shorter branches indicate a smaller number of genetic changes, suggesting a more recent common ancestor and closer evolutionary relationship.
- The **nodes** represent **common ancestors**. The tree shows that all the human glycophorin A sequences share a common ancestor.
- An **outgroup** is a **distantly related sequence** used to root the tree and provide a point of comparison.
- The first tree includes a sequence from a chimpanzee (*Pan troglodytes*) labeled **NP_002090.4**, which serves as a reference to show how the human sequences have evolved relative to a closely related species. As we can see from the first tree **Chimpanzee (*Pan troglodytes*)**, **Bonobo (*Pan paniscus*)** and **Gorilla (*Gorilla gorilla*)** shows lower branch length and thus serves as the closely related species.
- In the second tree AAC50057.1 is the closest to our query sequence.

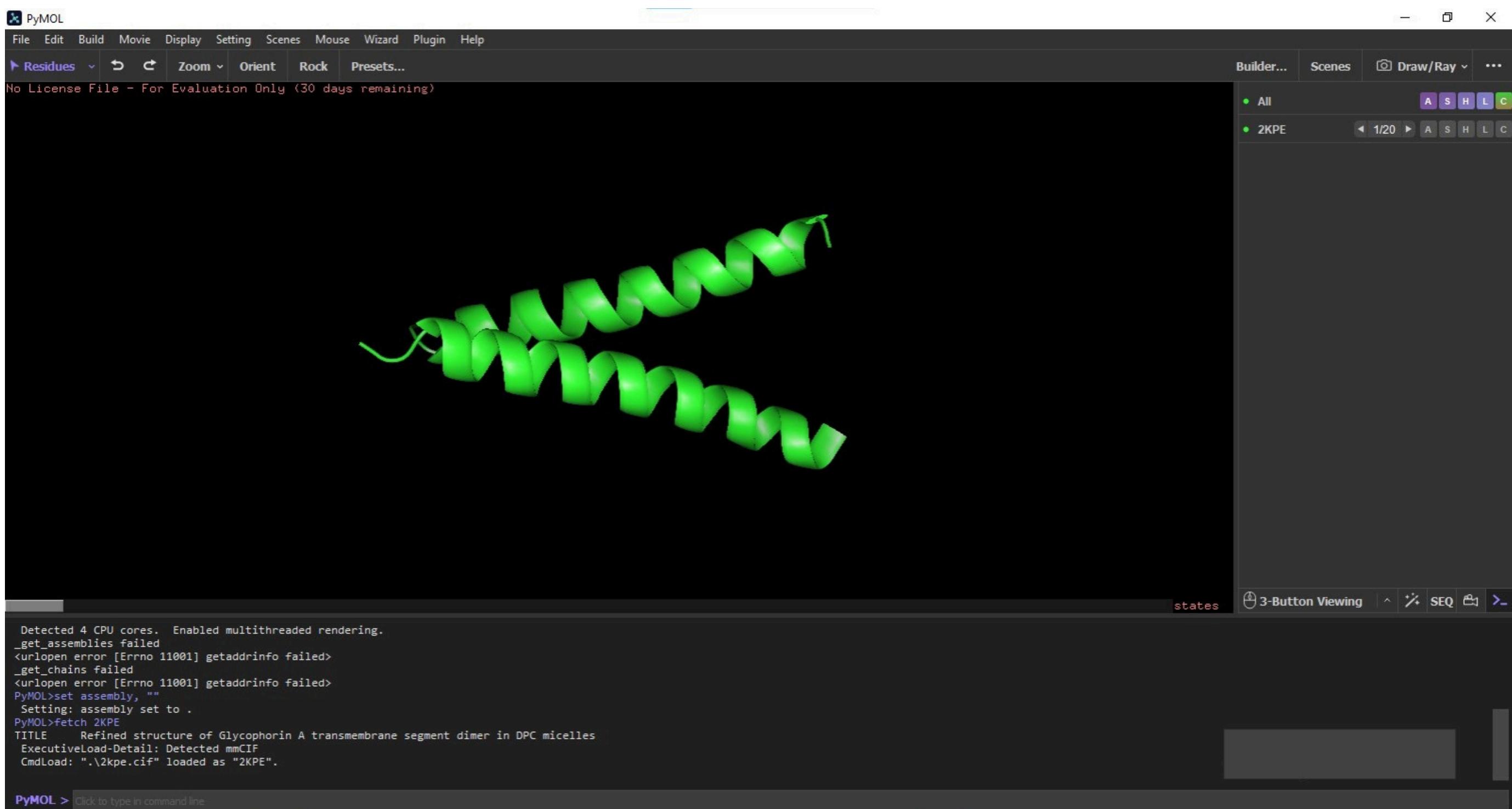


STRUCTURAL ALIGNMENT BY PyMol

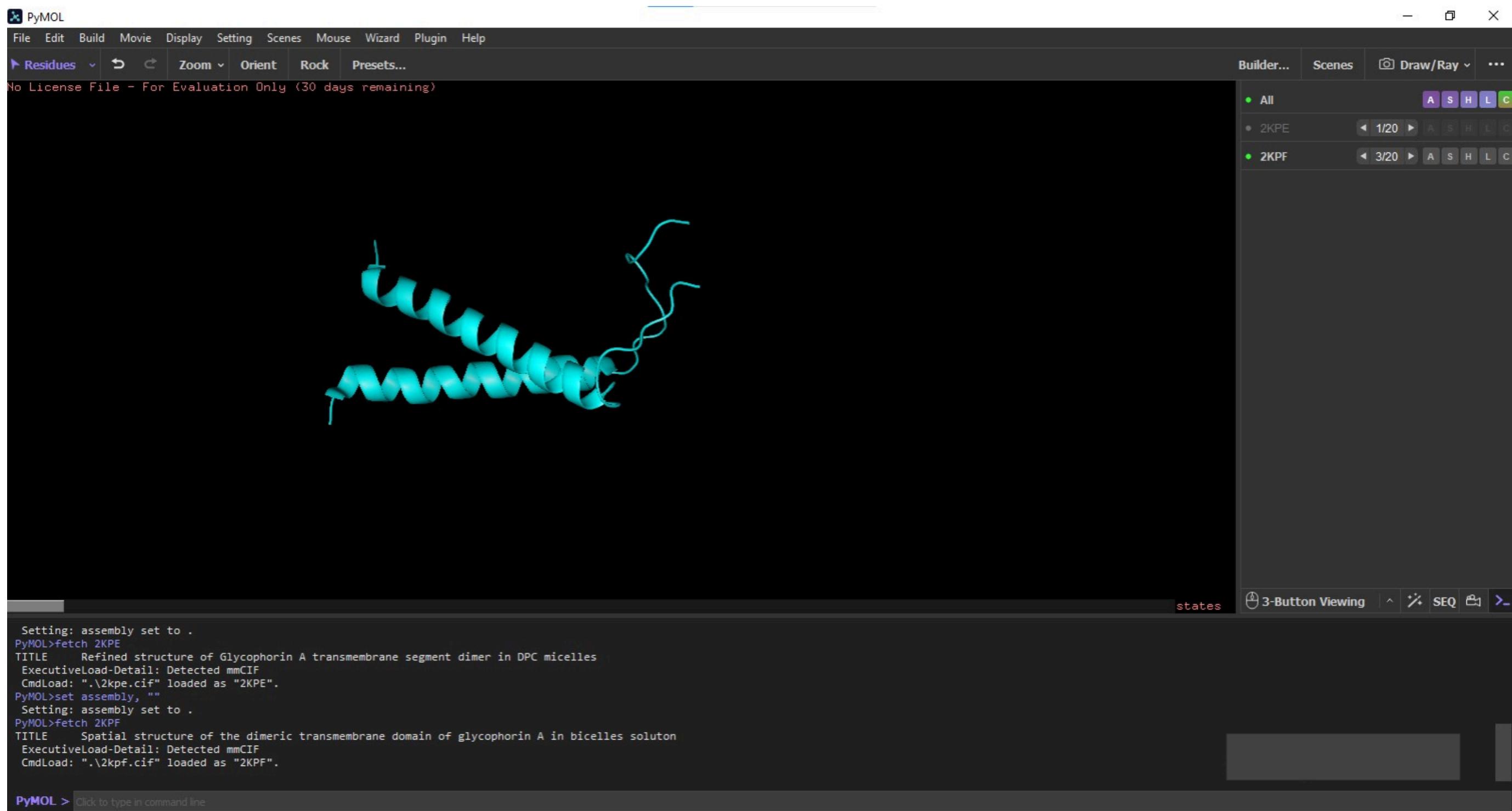
Process of comparing and superimposing 3D protein structures to identify similarities in their folds and spatial arrangements.

The structures of the Model & the Template are aligned in PyMOL.

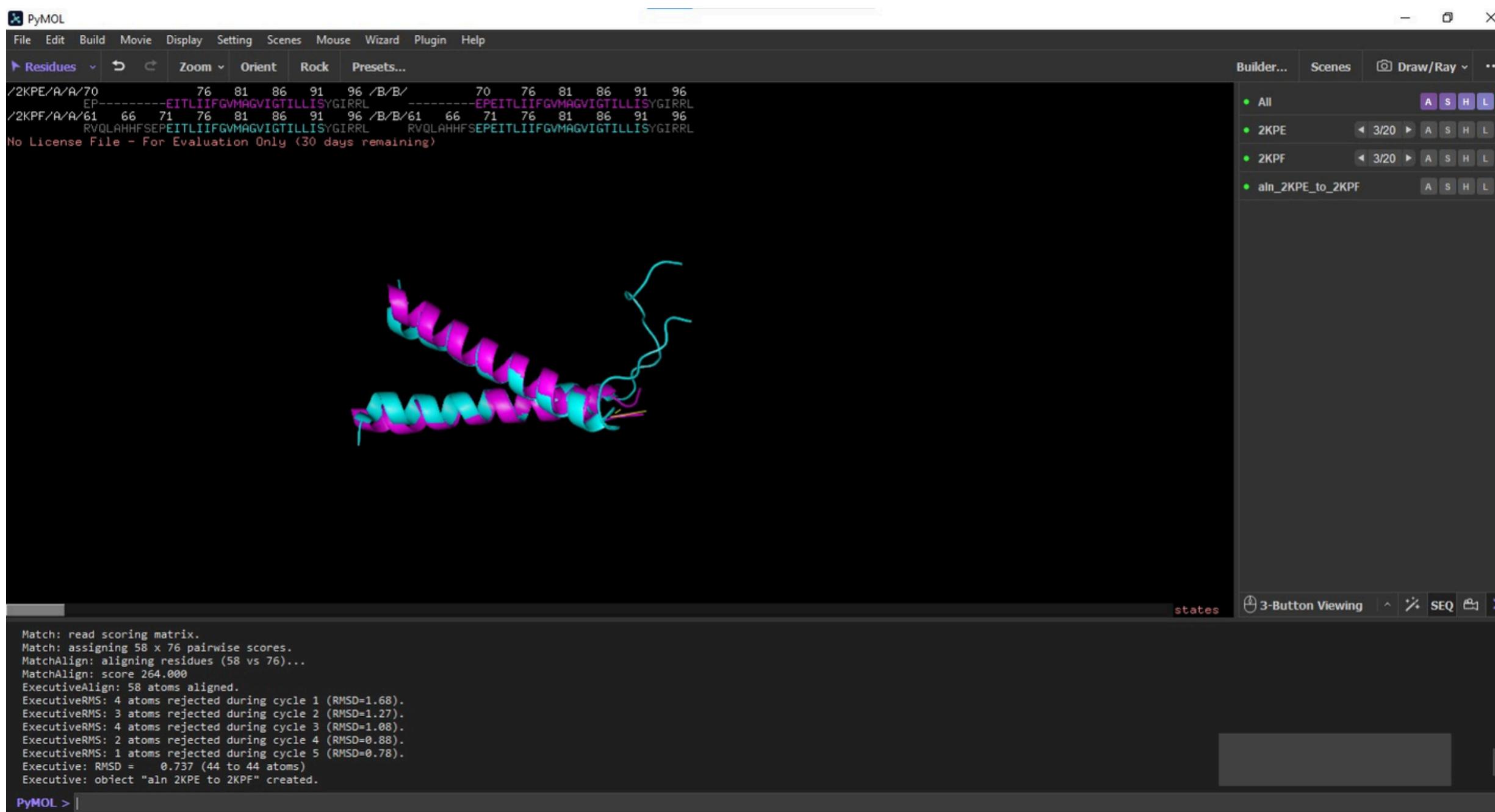
1. The .pdb file of the model was opened in PyMOL.



2. On the same PyMOL window, the .pdb file of the structure that is to be superimposed is opened.



3. The Model was aligned to the Template by aligning the “2KPE” object to “2KPF”. Both the structures are derived by NMR. These structures have multiple models. Here only one such model is considered for both the structures and then superimposed.



4. RMSD value of the alignment was noted, and the aligned molecules were exported with a white background after removing the solvent molecules (H_2O).

The RMSD value is noted to be 0.737 (below 1.5 \AA) which represents a high degree of similarity between the two structures.

```
Match: read scoring matrix.  
Match: assigning 58 x 76 pairwise scores.  
MatchAlign: aligning residues (58 vs 76)...  
MatchAlign: score 264.000  
ExecutiveAlign: 58 atoms aligned.  
ExecutiveRMS: 4 atoms rejected during cycle 1 (RMSD=1.68).  
ExecutiveRMS: 3 atoms rejected during cycle 2 (RMSD=1.27).  
ExecutiveRMS: 4 atoms rejected during cycle 3 (RMSD=1.08).  
ExecutiveRMS: 2 atoms rejected during cycle 4 (RMSD=0.88).  
ExecutiveRMS: 1 atoms rejected during cycle 5 (RMSD=0.78).  
Executive: RMSD = 0.737 (44 to 44 atoms)  
Executive: object "aln_2KPE_to_2KPF" created.
```

CONCLUSION

The following conclusions can be drawn for the protein of interest from this project:

- Detailed information about the **Glycophorin-A** from ***Homo sapiens***, including the data about the **Source Organism**, **Subcellular Localization**, **Structure**, **Sequence**, **Family**, **Domain**, **Function** and **Post-Translational Modification/Processing** was obtained from **UniProt**, **InterPro** and **Pfam**.
- The complete Protein sequence was subjected to a **BLAST search** at **NCBI-BLAST** and similar proteins based on sequence were found.
- The identification of both conserved and non-conserved residues was accomplished through **Multiple Sequence Alignment(MSA)** involving the entire protein sequence, alongside sequences obtained via **BLAST Search** at **MultAlin**.
- A phylogenetic tree was constructed using **ClustalW**. **NP_002090.4** and **AAC50057.1** were found to be evolutionary close to the protein of interest.
- The experimentally determined structure of the protein and our protein of interest were aligned in **PyMOL** and the **RMSD** value appeared to be **0.737 Å** (less than 1.5 Å) indicating high degree of similarity between the template and the structure chosen.

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to Dr. Jhimli Dasgupta, our respected and dedicated Bioinformatics professor, for her exceptional mentorship and unwavering support throughout the course of this project. Her insightful guidance, patience, and encouragement played a pivotal role in shaping our understanding of complex bioinformatics concepts. Under her expert supervision, we were able to effectively explore, understand, and apply a diverse range of bioinformatics tools and methodologies, which were instrumental in the successful execution and completion of our project objectives.

We would also like to express our special thanks to the numerous databases, websites and software tools that we relied on during the course of our work. These resources provided critical data, analytical frameworks, and visualization tools that significantly contributed to our research, data analysis, and the overall development of our project.

REFERENCES

- Tomita M, Furthmayr H, Marchesi VT. Biochemistry. Primary structure of human erythrocyte glycophorin A. Isolation and characterization of peptides and complete amino acid sequence. 1978 Oct 31;17(22):4756-70. doi: 10.1021/bi00615a025. PMID: 728384
- M Tomita, V T Marchesi. Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. DOI: 10.1073/pnas.72.8.2964 . PMID: 1059087 PMCID: PMC432899
- Ochoa D, Jarnuczak AF, Viéitez C, Gehre M, Soucheray M, Mateus A, Kleefeldt AA, Hill A, Garcia-Alonso L, Stein F, Krogan NJ, Savitski MM, Swaney DL, Vizcaíno JA, Noh KM, Beltrao P, The functional landscape of the human phosphoproteome. Nat Biotechnol, 38(3):365-373(2020) [pubmed]
- A Pisano, J W Redmond, K L Williams, A A Gooley. Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human Glycophorin-A . PMID: 8286855 DOI: 10.1093/glycob/3.5.429
- Furthmayr H. Structural analysis of a membrane glycoprotein: glycophorin A. J Supramol Struct. 1977;7(1):121-34. doi: 10.1002/jss.400070111. PMID: 604696.
- Kahane I, Gitler C. Red cell membrane glycophorin labeling from within the lipid bilayer. Science. 1978 Jul 28;201(4353):351-2. doi: 10.1126/science.663661. PMID: 663661

- Databases and Tools:

- 1.The UniProt Consortium.UniProt: the Universal Protein Knowledgebase in 2023,Nucleic Acids Res.51:D523–D531 (2023)
- 2.Basic Local Alignment Search Tool (BLAST). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2023.11.03]. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- 3.Pfam: The protein families database in 2021 J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman Nucleic Acids Research (2020) doi: 10.1093/nar/gkaa913
- 4.Edgar, R.C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5):1792-1797, PMID: 15034147, DOI: 10.1093/nar/gkh340
- 5.The PyMOL Molecular Graphics System, Version 2.5.5 Schrödinger, LLC.