Structural Bioinformatics Predicts Large Intrinsically Disordered Regions of Erythrocyte Binding-Like Proteins of Plasmodium sp.: Functional Implications

Arita Acharjee

DH-6/24, Street No. 0317, Action Area- ID, International Institute of Innovation and Technology, New Town, Kolkata- 700156.

Abstract

The Erythrocyte Binding Ligand (EBL) is an important protein of Plasmodium sp. required by the parasite for successful invasion into the red blood cells causing malaria. EBL is a potential anti-malarial drug target. Despite its importance only a portion of its structure i.e., Duffy Binding Domain (DBD), is solved. Thus, the remaining regions necessitate attention and probing. In this study, it is reported that EBL protein consists of massive unstructured regions in between two compact structured portions, viz., the DBD and a C-terminus trans-membrane domain as predicted from our bioinformatics studies on Plasmodium vivax. Our study reveals that its major parts of the unstructured portions consist of Intrinsically Disordered Regions (IDR). Since IDRs have high rates of plasticity, the hypothesis is that it allows large molecular movements between the two structured portions of EBL helping the DBD to find its target receptor. Here, a schematic of the overall structural organization of EBL and its functional model is proposed. Thus, our study signifies possible importance of hitherto unexplored 'disordered'-ness in EBLs in mediating pathogenesis in Plasmodium sp.

Key words: Plasmodium, Malaria, EBL, EBA, IDP, IDR, drug target, plasticity, evolution.

1. Introduction

Malaria is a fatal disease caused by the infection and proliferation of the protozoan parasite Plasmodium species, carried by arthropoda vector. In tropical countries numerous people die due to lack of effective anti-malarial drugs and drug resistance. Malaria transmission occurs widely in tropical climates such as large areas of Africa and South Asia and parts of Central and South America, the Caribbean, Southeast Asia, the Middle East and Oceania. Hence, to prevent malarial epidemic, we should focus on the mechanism of parasitaemia during malaria as well as the role of different proteins associated with this the disease. We can address how the associated proteins can be modified or targeted to influence the parasite and its rate of parasitaemia in mediating Malaria as the global number of deaths due to malaria is a growing concern (1).

When an infected Anopheles mosquito bites a human, the sporozoites are injected into the human blood stream and migrate to liver where it passes through kupffer cells and invades into human hepatocytes and gets to converted tmerozoite (a small, polar, pear

shaped cell). This is then released into the blood and invades erythrocytes. Here, the invasion ligands play a crucial role, identifying whose properties can be a significant approach to deal with malaria (2). These invasion ligands are from various families, such as- EBL, RBL etc., and are secreted from microneme and the neck of the rhoptries to the apical end of the merozoite. As these invasion ligands interact with the RBC-receptors, these can be the target-site of various potential blood-stage vaccines. Interaction or binding between these invasion ligands and RBC-receptors is very specific and some interaction ligands and their corresponding receptors are tabulated in Table 1.

Table 1: Invasion ligands and their corresponding RBC receptor(s) in *Plasmodium* species. The "question marks" in Table 1 of the manuscript indicate that the receptors corresponding to the specific ligands are yet to be identified and/or are not characterized yet

Invasion ligands of *Plasmodium* species	**RBC receptors**
EBA-175	Glycophorin-A
EBA-140	Glycophorin-C
EBA-181	?
EBL-1	Glycophorin-B
Rh5	Basigin
Rh4	CR-1
Rh-2a	?
Rh-2b	?
Rh-1	?
AMA1	RON-2
RBP1	?
RBP2	?
RBP-homologues	?

The Erythrocyte Binding Ligand / Antigen; also known as Duffy Binding domain (DBD) is one of the foremost and important invasive erythrocyte binding ligands used by the merozoites which indicate that it is a tightly controlled and pertinently crucial process involving specific and multiple receptor-ligand interactions between host and parasite molecules, making it a good tool and target for malarial drugs. Till now, five genes have been

identified those encode EBL proteins, including, Erythrocyte Binding Antigen-175 (EBA-175, MAL7P1.176), (3), EBA-140 (JESEBL, PFA0125C), (4, 5), EBA-181 (BAEBL, MAL13P1.60), (5-8), EBL-1 (PF13_0115); (9) and EBA-165 (PFD1155w) (10). It is also reported that the DBD forms the basis of cytoadherance as well as invasion into the erythrocytes in malaria as a forerunner (11). EBL protein consists of 'F' region (region-II) that contains two domains, F1 and F2, having receptor binding capacity which is noticed almost in all Plasmodium sp. This is the portion which is well characterized and mostly structured (12-14). Binding of EBA-175 and EBA-140 of P. falciparum with their corresponding RBC-receptor glycophorin-A (15) and glycophorin-C (16), respectively are quite sialic acid dependent. Region-II in EBA-175 is polymorphic, whereas it is less polymorphic in case of EBA-140 and EBA-181.This slight polymorphism in binding region of EBA-140 and EBA-181 affects the receptor binding affinity but is unable to disturb the specificity of them (17). N- terminal end of the EBA-175 contains cysteine-rich DBL-domain that form dimer in a "handshake" manner (18) while interacting with the sialic acid dependent RBC-receptors. The peptide backbone is also important for the specificity of this receptor-ligand binding. When a couple of DBL domains dimerize, the interdomain channels are involved which have a strong receptor-binding potential (19). The region-II of the DBL-dimer in P. falciparum EBA-175 interacts with both peptide backbone as well as sialic acid side chains. The cytoplasmic-tail of DBL-domain is attached with the actin-myosin complex through the Thrombospondin Anonymous Repest Protein (TRAP) (20). Via proteome analyses, much disordered percentage has been accounted for in Plasmodium sp. and EBL protein, though specific zones were not highlighted (21). So, while correlating the entire sequence of EBL protein and their crystal structures as obtained from RCSB PDB (22), a large stretch of sequence was found to be unaccounted for, in the structure. To uncover the disparity, it is intended to identify the stability and percentage of disorder of the unstructured portion and specify their intrinsic properties.

In any organism, a large portion of the proteome consists of polypeptide segments, which have no well-defined or fixed secondary or tertiary structures (23-25) but these segments are flexible (26) and movable (27). These protein segments are known as Intrinsically Disordered Regions (IDRs). And the proteins, which consist of disordered sequences entirely, are known as IDPs. They can do these functions via structural modification as well as allosteric modulation (28). Near about 44% of human protein-coding genes contain disordered regions having >30 amino acids in length (29) and 6.4% of all protein coding genes do not even have any functional annotation. Some common in vitro and in vivo IDP rich proteins are: MAP2, Myelin basic proteins, α-synuclein, α-fetoprotein, tau proteins, prion proteins etc.

IDPs lack bulky hydrophobic amino acids (i.e. low hydrophobicity) (30), hence they are unable to form hydrophobic core that can construct a structured domain, as a result of which they have a quite different functionality unlike the structured proteins (31,32). These equally charged residues cause electrostatic repulsion that leads to a high disorder in their structure (33). The IDPs are mostly Proline rich (34) which may have a tendency to undergo cis-trans isomerization, that may inhibit the protein to form a compact structure. This dynamicity makes the structure ''STABLE'' thermodynamically, during post-translational modification (PTM), such as phosphorylation, deamination and acetylation. Hence the net charge of the protein is changed that affects the structural property and binding affinity of IDPs towards another protein (35-38). In bound state, the IDPs have different conformational ensembles with a particular entropy. Hence, the free energy of the structure varies, thereby changing the conformations. Weak interactions among the charged amino acid residues cause transient ordering leading to fuzzy complex (39,40). IDPs have short linear motifs of 3-10 amino acid residues that allow the interaction of IDPs with the structured domains of another protein or, biomolecules (DNA, RNA etc). PEST

motifs are such type of motifs which are rich in proline (P), glutamic acid (E), serine (S), threonine (T) and first observed in eukaryotic intracellular proteins, those of whom are rapidly degraded. Upon binding, these flexible linear motifs cause a transient secondary structure, termed as PreSMOS (pre-structured motifs) (41). From several experiments it is observed that many viruses mimic the linear motifs and hence through a comparative mechanism those viruses can easily prevent the linear motifs from their target-binding site leading to different diseases. Hence research on this topic may open up several avenues to understand and investigate the cause and remedy of the diseases.

IDPs are also known to adapt many different structures by creating conformational ensembles and can interact with biomolecules. So, their functions entirely depend on their intrinsic plasticity. Flexibility of IDPs somewhat facilitate the conformational changes so that they can bind with the modified enzymes and their receptors (42). It is observed that IDP-rich proteins participate in cell signaling, transcription and chromatin remodeling. IDPs can act as linkers between the structured domains allowing them to rotate or twist freely through the protein domains. Chaperone proteins that help the RNA and protein molecules to reach their folded state to perform specific functions also (43,44) consist of IDPs. Effector proteins, that interact with another protein to regulate its activity has IDRs as its component (45, 46). The dynamicity of IDPs contributes to its unique quality and entropy plays an important role in imparting these dynamic characteristics an edge.

IDPs are responsible for a number of diseases (47). Due to the structural flexibility of IDPs, they can aggregate with each other in a random manner when their concentration is high, resulting in many synucleinopathies as well as toxicity and can lead to cancer or cardiovascular diseases. Many oncogenes consist of IDPs e.g., p53 and BRCA that can cause cancer. Disordered segments of proteins may be involved in some non-functional interactions and thus can create an imbalance in signal pathway (48). Disordered proteins are encoded by over expressed harmful genes (dosage sensitive genes) (49). Availability of IDPs can be controlled by multiple mechanisms at different stages during gene expression i.e., from transcription to protein degradation (50). Thus regulation causes the availability of IDP in an appropriate level. Considering the cells' basic defense mechanism many drugs can be developed that can block the binding site of the toxic substrates and inhibit them and hence can prevent diseases. IDPs thus can be focused as potential drug targets, thereby increasing the possibility of EBL, RBP and other related proteins as a candidate for vaccine development.

EBL has already been established as one of the primary drug targets for malaria but to target the IDRs of the EBL protein in itself will be able to open a new vista to probe into the dynamic properties of IDRs and how the adjacent structured portions facilitates its plasticity. Using PONDR, the percentage of disorder in a protein sequence from the VLXT value is identified. If the VLXT value for each amino acid is greater than equal to 0.5, it will be strongly disordered. The greater the VLXT value the greater is the plasticity of the protein given. Other than the perceived notion of having fully structured proteins, it has to be kept in mind that unstructured and intrinsically disordered proteins (IDP) are a continuum of the entire paradigm. The unstructured portion dealt out to be partially structured whereas the rest bulk was found to be intrinsically disordered which made the investigation more interesting as IDPs and their regulation can make EBL a primary tool and target for anti-malarial drugs in this crucial hour. Alongside similar family proteins involved in parasitaemia were investigated to uncover their plastic properties, if any, and find the avenue for further routes of development of the proteins as potent drug targets.

This study itself will help to explore the evolutionary link and model EBL as one of the most vital drug targets for malaria, alongside other parasitaemia mediating proteins in correlation to other blood borne disease parasites.

2. Materials and Methods

Softwares used for prediction and hypotheses

NCBI (51) had been used to get the genome sequences as well as the amino acid sequences of the EBL proteins of different Plasmodium species. To construct the automated and template based homology models (taking P. vivax EBL, PDB ID: 3RRC and P. falciparum EBL, PDB ID: 4GF2 as template) of these EBL proteins of the Plasmodium species I have used SWISS MODEL (52). CLUSTALW (53) had been used to perform the Multiple Sequence Alignment (MSA) among the EBL of those Plasmodium species to find out the similarity and identity percentage among them, and also the conserved amino acid residues / sequences / domains in them. PSIPRED (54) software was used to identify not only the disordered regions in the EBL protein sequences but to predict the secondary structure also, FFpred and Memstat tell us about the probable functional characteristics and the transmembrane regions present in the last structured portion of the EBL proteins respectively. PONDR (55) had also been used to ensure the presence of the Intrinsically Disordered Portions (IDPs) in the EBL proteins, in the VLXT method the cut off value for disorder prediction is 40 which implies that when the "overall percent disorder" value is greater than or equals to 40 then it can be considered as IDP. ProtCalC (56) had been used to identify the net charge of the amino acids present in the specific sequences of those proteins (the sequences may be total sequence, IDPs, last structured portions of EBL proteins), whereas Protparam Expasy (57) had been used to identify the percentage of different amino acids present in a specific sequence of the EBL protein as well as to identify the molecular weight of that specific portion of the protein. Uniprot (58) had been used to find out the Uniprot ID of EBL protein in a specific Plasmodium species. SlimPred (59) gave us the coverage of the ordered and disordered portions present in the EBL proteins in different Plasmodium species pictorially and also mentioned the specific motifs present in them. ProtScale Expasy (60) had been used to determine the hydropathy plots in the Kyte-Doolittle method, actually it presented a position of amino acids versus score curve, in which the negative value/score implied lower hydrophobicity and the positive value indicated greater hydrophobicity, thereby indicative of transmembrane regions. ANCHOR (61) had been used to find out the binding sites present in the IDP regions of EBL protein. RAMPAGE (62) gave us the Ramachandran plots of EBL proteins of different Plasmodium species and showed their allowed region.

3. Results

Among all ~250 Plasmodium species, only sixteen Plasmodium species had been found to have genome sequences. As EBL is not an extensively studied protein, other than in case of P. falciparum EBA-140, EBA-175 and partially in case of EBA-181 and MAEBL (a partial analyses on their structure and format has been done), the focus is this EBL protein present in all the known Plasmodium species having genome sequences, which are determinants to infect and spread the disease in various species. Thorough bioinformatic and structural studies were conducted on them, in homology to the already known structures of P. falciparum EBL variants. A chart of the aforesaid Plasmodium species and the number of amino acids in their EBL proteins and corresponding molecular weight has been presented in Table 2.

However, only two models so far (one each as representative of a species) have been known for the structure of the EBL protein, in case of P. falciparum EBA-140 (PDB ID: 4GF2) and P. vivax EBL protein (PDB ID: 3RRC) based on which the homology modelling of the rest species are done, in relation to both. Moreover, a comparative chart is presented in Table 3, which depicts how many amino acids are actually involved in the structure formation.

Moreover, it was also noticed that all the EBL proteins present in various Plasmodium species maintain a greater similarity% / identity% with the P. vivax EBL (PDB ID: 3RRC) than that of P. falciparum EBA-140 (PDB ID: 4GF2) while doing a comparative study using CLUSTAL-W Pairwise Alignment, as shown in Table 4.

Table 2: Chart of EBL proteins of different Plasmodium species and their corresponding molecular weight.

Plasmodium sp.	Total number of amino acids present in EBL	Molecular Weight
P. vivax	1061	118371.88
P. falciparum (EBA-140)	1210	140596.06
P. ovale	858	99394.16
P. vinckei	783	90698.03
P. yoelii	852	98440.19
P. inui	872	100999.26
P. berghei	827	95716.60
P. malariae	819	95068.72
P. knowlesi	1073	120784.14
P. reichenowi	683	78807.05
P. cynomolgi	1050	119534.78
P. coatneyi	945	105803.60
P. gaboni	584	68128.55
P. chabaudi	785	90688.96
P. fieldi	1004	113809.75
P. simiovale	988	111810.00

Table 3: Chart of the total number of amino acid residues present in the EBL proteins and the range of amino acids involved in the homology model with P. vivax and P. falciparum

Plasmodia species	Total number of amino acids in EBL	Amino acids involved in homology w.r.t *P. vivax*	Amino acids involved in homology w.r.t *P. falciparum*
P. ovale	858	227-516	232-514
P. vinckei	783	115-412	111-414
P. yoelli	852	132-423	105-424
P. inui	872	55-130	55-499
P. berghei	827	132-429	158-426
P. malariae	819	240-502	215-504
P. knowlesi	1073	208-505	-
P. reichenowi	683	156-443	155-430
P. cynomolgi	1050	201-498	-
P. coatneyi	945	180-477	-
P. gaboni	584	17-281	46-280
P. chabaudi	785	120-411	109-412
P. fieldi	1004	201-468	201-465
P. simiovale	988	197-452	199-450

Table 4: Comparisons among the percentages of similarities/identities of EBL proteins in different Plasmodium sp. with respect to the P. vivax EBL and P. falciparum EBA-140

Plasmodium sp.	Similarity% / Identity% with respect to *P. vivax* EBL	Similarity% / Identity% with respect to *P. falciparum* EBA-140
P. ovale	27.9/18.3	24.4/14.8
P. vinckei	27.1/18.1	27.9/16.8
P. yoelii	26.5/16.0	19.3/11.5
P. inui	33.8/26.4	28.5/15.9
P. berghei	26.7/16.8	18.9/11.2
P. malariae	27.5/17.1	24.4/14.8
P. knowlesi	31.8/26.2	18.0/10.6
P. reichenowi	29.3/17.3	27.8/16.6
P. cynomolgi	38.9/35.3	18.1/11.2
P. coatneyi	35.6/27.7	26.7/16.0
P. gaboni	32.8/19.0	19.7/11.9
P. chabaudi	29.4/17.7	18.0/10.0
P. fieldi	79.3/72.6	33.0/21.5
P. simiovale	77.7/70.9	32.7/19.9

As the EBL proteins in most of the Plasmodium species have a greater similarity with that of P. vivax, the result is being described with respect to the EBL protein of P. vivax.

From Table 3, it is clear that the entire amino acid sequence of EBL cannot form protein structure rather a long range of amino acids in the middle portion form compact protein structure leaving some anterior and posterior regions unstructured. Using SLIMPRED (Supplementary Figure 1), the prediction of the structure of the EBL protein is been made, where a small unstructured portion (U-1) is present followed by a structured portion (S-1) and further an unstructured region (U-2) is followed by a structured portion (S-2) evolving a short unstructured tail and the trend was seen to be similar in all other Plasmodium species whose EBL protein has not been investigated, so far.

U-1 is very small in all the Plasmodium species except P. vivax (up to ~200 amino acids). The next S-1 (first structured portion) contains up to ~525 amino acids, U-2 contains around ~ 825 amino acids and the S-2 (last structured portion) is up to ~1050 amino acids and the rest amino acids up to ~1061 forms the short tail. Based on the studies, a representative schematic has been drawn of the primary structural organization of full length EBL of P. vivax (see Figure 1A) (the segments as indicated, though proportional, are not in scale).

The first structured portion, S-1 (200-525 amino acids) of P. vivax EBL forms a compact protein structure (PDB ID: 3RRC). It is concentrated on the unsolved U-2 and S-2 regions.

3.1. Characterization of last unstructured or disordered region, U-2

Noticing that a massive part is actually unstructured raised a question as to what could have been the properties of these specific unstructured regions. This led us to study whether the unstructured regions can

be intrinsically disordered (IDRs) or not and hence explore their salient features.

3.1.1. To identify and ascertain whether the claimed IDP regions actually are intrinsically disordered or not

For further affirmation and validation, four routes were taken.

Presence of disorder

Subjecting the sequences to PONDR and verifying it with the DISOPRED (PSIPRED series), softwares to identify not only the disordered regions in the EBL protein sequences but to predict the secondary structure also, led us to believe that our suspect regarding the unstructured regions being Intrinsically Disordered may be actually true (Figure 1B and 1C). With the disordered percentage overall increasing more than 40%, a standard threshold considered in bioinformatics study to ensure whether the protein or peptide sequence under study has a classic intrinsic nature of disorder or not. More the value, (>40%), more tends to be the disordered dynamics.

PEST motif identification

The PEST motif usually is an identification marker for the commonly known

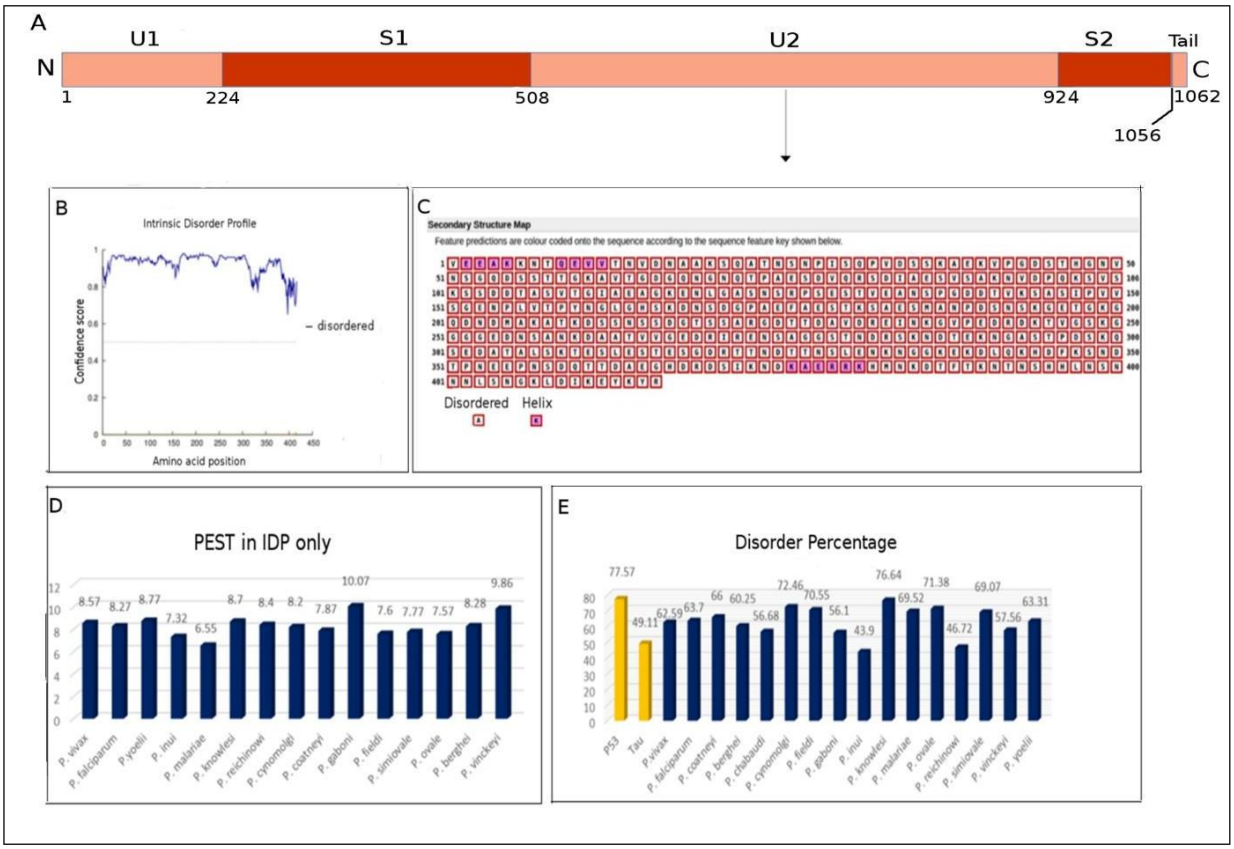

Figure 1: (A) Schematic of primary structural organisation of full length EBL of P. vivax. (B) PONDR (VLXT) of U-2 portion of P. vivax EBL indicates that there is 97.35% disorder. (C) PSIPRED result of U-2 portion of P. vivax EBL indicates that there is lots of disorder (red boxes). (D) Percentage of PEST sequence in U-2 portion of EBL in Plasmodium species indicates that U-2 is PEST-rich region. (E) Percentage of disorder in commonly known IDPs (Tau, P53- in yellow) and EBL of Plasmodium sp. (in blue) shows that EBL U-2 is a disordered region.

Intrinsically Disordered Proteins (IDPs). Here, in relevance to the commonly known IDPs, a chart for the percentage of PEST motif present in EBL proteins of the Plasmodium sp. (known sequences) is presented (Figure 1D). The identification

marker is representative of the fact that with increasing PEST motif or in other words, more abundance of the amino acids Proline (P), Glutamate (E), Serine (S) and Threonine (T), irrespective of them being in sequence or not helps to determine whether the propensity of a particular protein is to attain intrinsically disordered dynamics or not.

Comparative graph analyses

A comparative plot as depicted in Figure 2 shows the similarity in the percentage of disorder in normally known IDPs (such as- Tau, P53 etc.) as well as the EBL proteins in our Plasmodium species under study (Figure 1E). With this graph, an analogy or rather a comparison has been done to check how similar this protein under consideration is, with respect to other known intrinsically disordered regions. The graph comparison clearly depicts that it is comparable that all the proteins in different known Plasmodium species shows striking similarity to that of the control proteins, known widely as intrinsically disordered.

Ramachandran Plot

The IDPs are rich in poly proline II motifs and they have their significant allowed regions in Ramachandran Plot for commonly known IDPs. Ramachandran plots for the unstructured regions, suspected as IDPs, were plotted and as shown in Figure 3, the allowed regions for the commonly known IDPs and our unstructured regions of the Plasmodium species under study are similar, reflecting on the fact that the unstructured regions, as suspected are genuinely Intrinsically Disordered. To generate the Ramachandran plots of EBL proteins and to demarcate their allowed regions, RAMPAGE was used (Figure 2A).

3.1.2. Net charge of the amino acids present in the U-2 region of EBL

The net charge was measured and compared to have an idea of the condition of the protein in physiological condition and to reflect on its hydrophilicity/hydrophobicity. This gives an indirect idea as to whether the protein remains in a bound/compact state (if hydrophobic) or tend to remain in soluble/in free access regions of the solvent (if hydrophilic). A comparative study of the net charge (ProtCalC.sourceforge.net) of the total amino acid sequence in the EBL protein and the amino acids present in the U-2 region indicates that the net charge of the U-2 portion is more bulky and consisting of negatively charged amino acids i.e. more hydrophilic (less hydrophobicity) than that of total sequence (Table 5).

Thus, as depicted, this less hydrophobicity of the U-2 region inhibits them to form a compact structure and in turn prevents to anchor properly with the membrane, it can be predicted that the reason behind them being unable to show trans-membranous or membrane-bound property.

3.1.3. Hydropathy plot

Hydropathy plots are good markers for identifying the hydrophobicity/hydrophilicity of a protein, depending on the score projected in a particular method. Hydropathy plots were carried out by Kyte-Doolittle method, a process that is indicative of reflecting the hydropathy context of a protein depending on the particular window size. For individual U-2 regions of the EBL proteins of the Plasmodium species, the above point was validated, referring to its hydrophilicity, which is hence proved to be water soluble in nature (see Figure 2B).

Table 5: Comparison of the net charge between the whole EBL sequence and the U-2 region

Plasmodium sp.	Net charge of the whole sequences	Net charge of U-2 region
P. vinckei	-19.0	-11.6
P. yoelii	-20.3	-18.6
P. inui	-21.3	-20.6
P. berghei	-19.8	-14.4
P. malariae	14.8	-4.3
P. knowlesi	-30.2	-41.3
P. reichenowi	-39.4	-41.0
P. cynomolgi	3.9	-5.1
P. gaboni	-44.5	-50.7
P. coatneyi	-26.8	-26.9
P. chabaudi	-20.1	-13.1
P. vivax	-6.2	-25.3
P. falciparum	-6.5	-28.8
P. fieldi	-17.8	-15.8
P. simiovale	-5.8	-16.3
P. ovale	4.0	-3.3

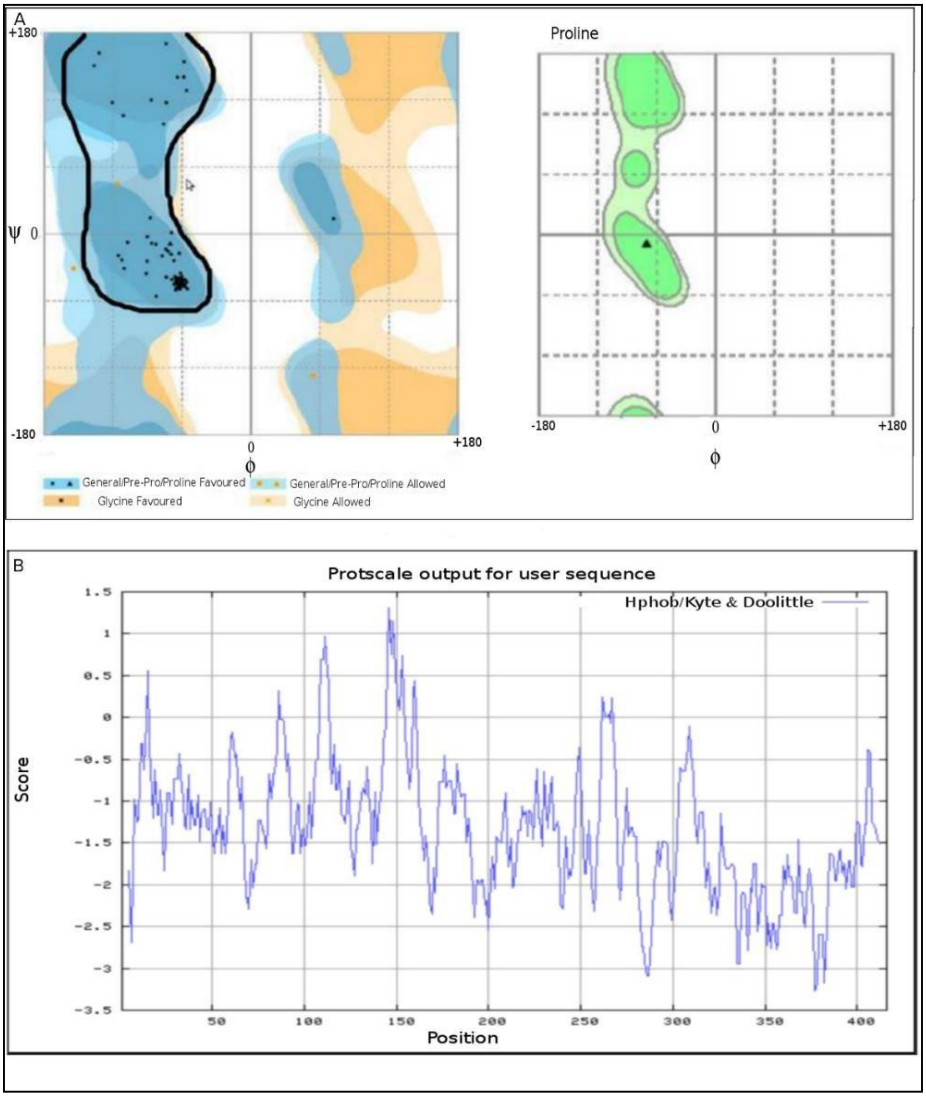

Figure 2: (A)Ramachandran Plot for U-2 region of P. vivax EBL indicates that U-2 is enriched in poly proline II motifs, signifying that U-2 is an Intrinsically Disordered Region (IDR). (B) Hydropathy Plot for the U-2 region of P. vivax EBL implies that U-2 is more hydrophilic (less hydrophobic) in nature.

3.1.4. Binding partner prediction of U-2 region

To find out in partial authenticity, the possible interacting mechanisms of the protein EBL, a software tool, ANCHOR, was used to find out the potential binding sites of the U-2 regions of the respective EBL proteins of each Plasmodium species. ANCHOR prediction of U-2 of P. vivax EBL thereby reveals the probable binding sites of the U-2 region (see Figure 3A).

3.1.5. Comparison between the U-2 regions of EBL protein of different Plasmodium species and prediction of conserved residues

The unstructured regions in proteins with intrinsically disordered nature may signify plasticity and thereby evolutionary competence. Thus, these unstructured regions are rich centers for extracting much more information on their evolutionary mechanism. Based on the sequences, a phylogenetic tree was constructed for the EBL proteins of the different Plasmodium species (see Figure 3B) and subsequently all those sequences were aligned on the basis of multiple sequence alignment in Clustal-W (see supplementary

Figure 2 The unique matter noticed in such an alignment is that the evolutionary tree constructed shows correlation with the similarly grouped IDPs of individual Plasmodium species. This indicates how IDPs may influence in the evolutionary competence of a particular protein, in this case, the EBL proteins of the Plasmodium group.

Figure 3B and Table 6 indicate quite strong evolutionary correlation to indicate the importance of EBL and its plasticity in terms of evolution.

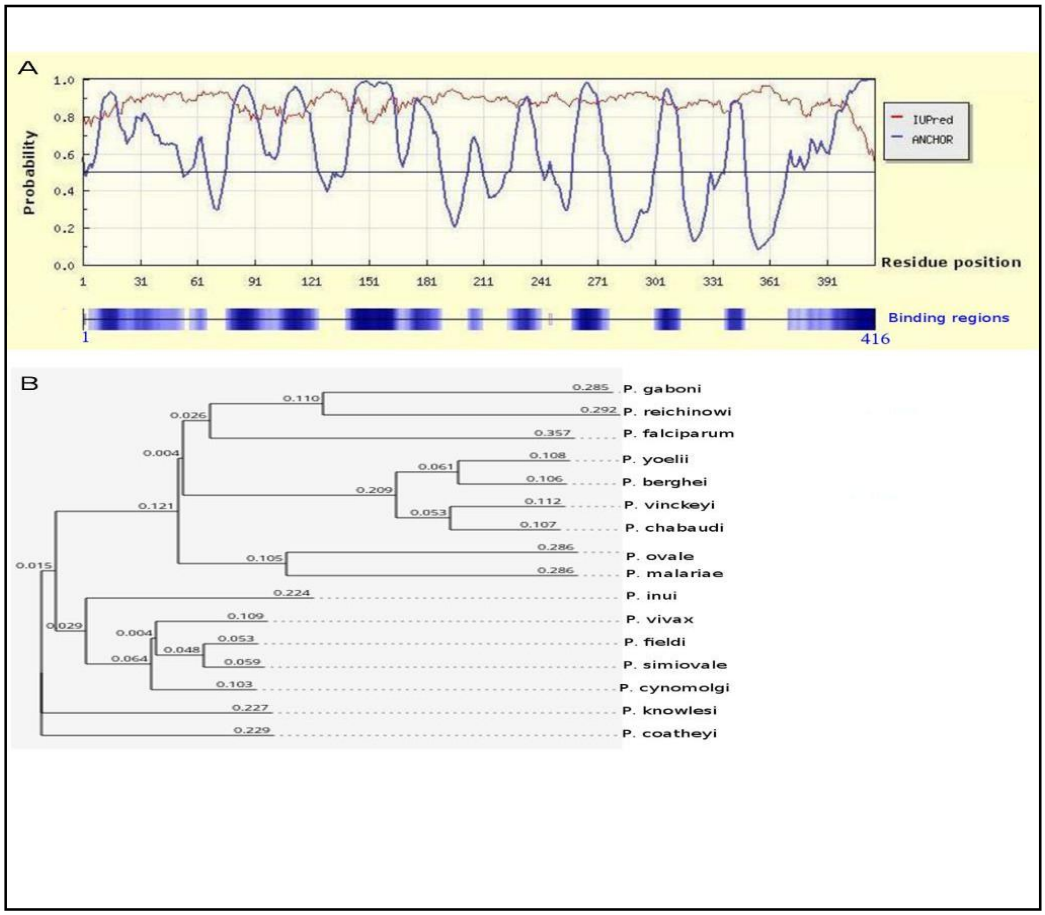

Figure 3: (A) ANCHOR prediction of U-2 of P. vivax EBL indicates the probable binding sites of the U-2 region. (B) Phylogenetic tree of second disordered portion U-2 of Plasmodium sp. from Clustal W.

Table 6: EBL- correlation with respect to P. vivax in terms of identity percentage

IDEN-TITY PERCENTAGE	*P. fieldi*	*P. simiovale*	*P. yoeliii*	*P. berghei*	*P. reichenowi*	*P. ovale*	*P. malariae*	*P. gabonii*	*P. chabaudi*	*P. inui*	*knowl*	*P. coatneyi*	*P. cynomolgi*	*P. vinckeyi*
16-17 %			√	√	√									

17-19 %						√	√	√	√					√
26-28 %										√	√	√	√	
30-36 %														
65-75 %	√	√												

3.1.6. Conserved amino acids in IDP

To look for the residues which actually contribute to the importance of forming these IDPs, other than the conservative motifs as identified earlier in the entire protein length, the multiple sequence alignment of the EBL protein of the Plasmodium species under study, reveals these conserved amino acids as single entities or in stretches which can be conveniently said to be playing a crucial role in forming these IDPs (supplementary Figure 3). The importance of the conserved residues can be implicated by conducting an Alanine Scan Mutagenesis, to specifically identify the importance of the particular residue or it being in a stretch of amino acids and further elucidating their structure-function correlation.

Therefore, the U-2 region can be termed as an Intrinsically Disordered Region (IDR) in the EBL protein of all Plasmodium species, by bioinformatic methods and the salient features observed can be pertinent to elucidate its functional aspects. It consists of a significant portion i.e., an average of ~38% of the total protein.

3.2. Characterization of the last ordered or structured portion, S-2

SLIMPRED result (Supplementary Figure 1) depicts a very important point that there is a structured portion followed by the U-2 region of the EBL protein in most of the Plasmodium species, termed as Last Structured/Ordered Region, S-2. In other words, it can be said that the IDR (U-2) lies in between two adjoining structured regions. The last structured part (S-2) on BLAST shows significant similarity to that of the P. falciparum EBA-175 for several Plasmodium species. This triggered interest in elucidating the structure of the last structured part to get better insights into its functional aspect and its possible association with the IDP regions.

3.2.1.1. PSIPRED Result: Prediction of ordered region (S-2)

The PSIPRED result, like before, of the S-2 portion (as indicated in Figure 4A) predicts that there is negligible amount of disorder (according to PONDR (VLXT method), the disorder is 14.39% i.e. <<40% in case of P. vivax) so S-2 can be considered as "ordered/structured portion" (Figure 4B and 4C). The same thumb rule was followed here as well in order to predict the disordered-ness of the protein sequence subjected.

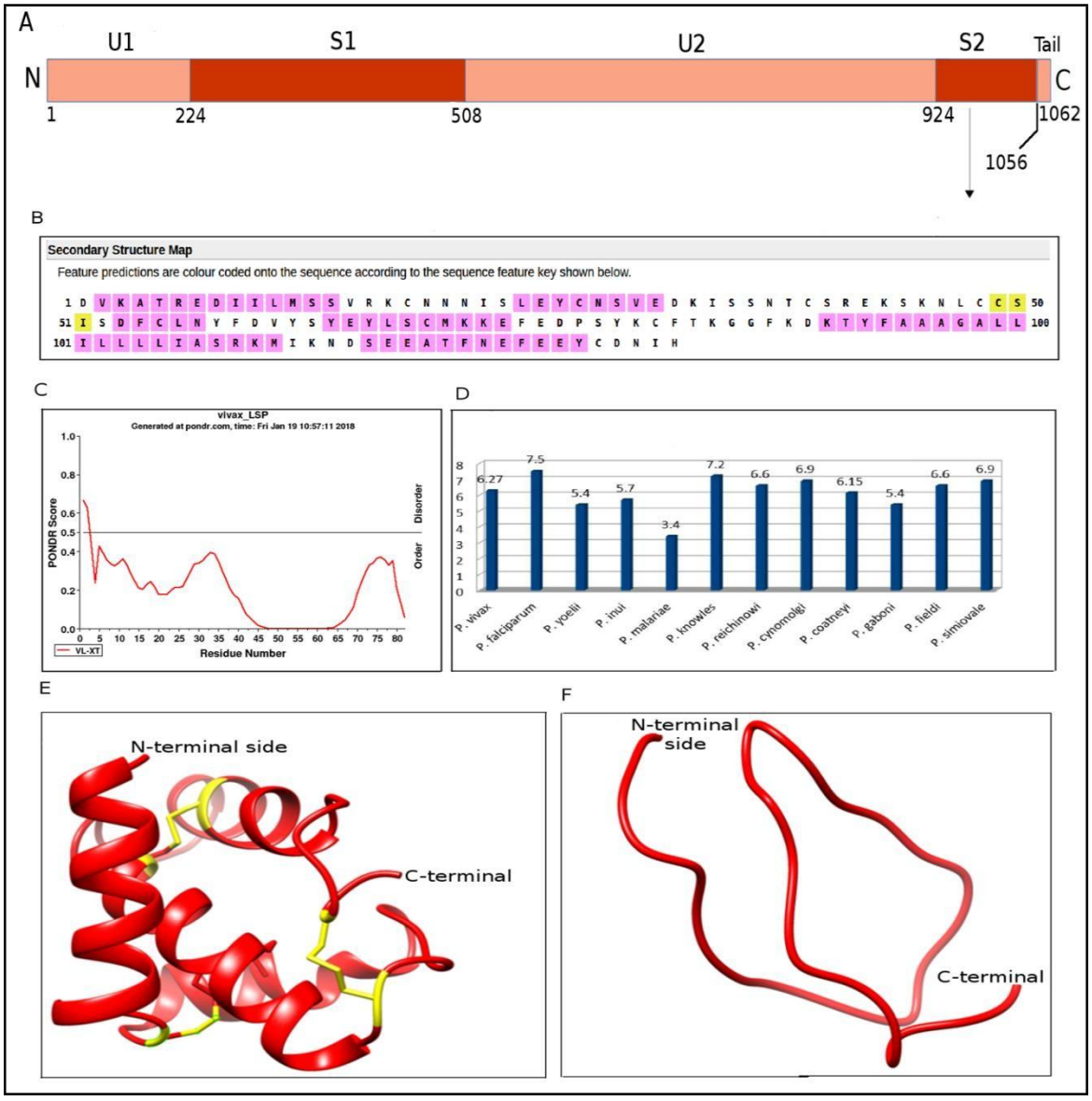

Figure 4: (A) Schematic of primary structural organisation of full length EBL of Plasmodium vivax. (B) PSIPRED result of the S-2 portion of P. vivax EBL indicates that the S-2 portion is ordered because the disorder is negligible amount (14.39%). (C) PONDR (VLXT method) of S-2 of P. vivax EBL agrees with the PONDR result. (D) Percentage of PEST sequence in the EBL S-2 region of P. vivax indicates that S-2 is PEST-deficient. (E): Last ordered portion (S-2) of P. vivax EBL containing four disulphide bonds (yellow bonds). (F): The relaxed thread-like structure of S-2 portion of P. reichenowi EBL protein.

3.2.1.2. PEST sequence prediction

A comparative study of PEST sequence present in the total EBL protein (Table 6), the IDR i.e., U-2 and in the last ordered portion S-2 showed that IDR portion is PEST-richer portion than that of S-2 which also proves that PEST is an identification marker of IDP regions. (Figure 4D). With less abundance of the individual amino acid residues of P, E, S and T, the propensity of the intrinsic disorder dynamics dropped to a huge extent in this case. The amino acid residues, thus suggests, more ordered packing of the S-2 region.

3.2.2. Structure-Function prediction

The last structured part of the EBL proteins of the Plasmodium species as modelled on the basis of homology with respect to EBA-175 of P. falciparum were studied in CHIMERA. The structure reveals a very interesting characteristic that the structured part of most of the Plasmodium species EBL are quite closely

knitted due to disulfide bonds, which if broken can dislodge the structure, as depicted in the Figure 4 E&F.

Using PSIPRED analysis, certain conclusions were drawn regarding the possible functions of this last structured region S-2 and its adjoining IDP regions U-2 as depicted in the modelled structures from its amino acid sequences. As in case of P. vivax, it was predicted via FFPred that this S-2 plays important roles in various functions such as regulation of metabolic processes, transport, cell communication, response to stimulus, signaling and signal transduction processes, developmental processes and others. From the FFPred and MemStat (to identify membrane binding properties of a region or the stretch of amino acids of a protein as incorporated as input) results, online functional prediction tools, it is also observed that this S-2 portion of all Plasmodium species (except P. malariae and P. reichenowi) have some trans-membrane property due to which, this can easily anchor with the membrane unlike that of IDRs which have no trans-membrane property (Figure 5). But in case of P. malariae and P. reichenowi, the S-2 portion is too short (consisting of a very narrow or short stretch of amino acid) to form structure and do not even have any disulfide bond to support the S-2 to be compact, the reason why they form relaxed thread like structures as comparatively depicted in Figures 4 E&F.

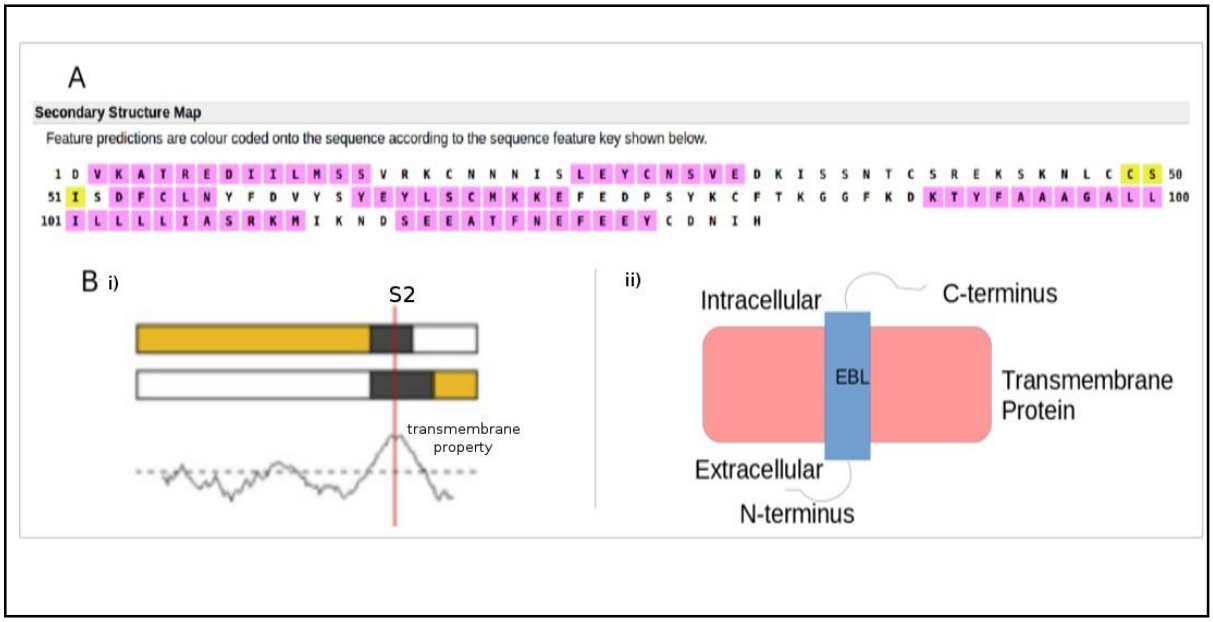

Figure 5: S-2 portion of P. vivax EBL is an ordered alignment as predicted from DISOPRED in (A), (B)(i) shows that it has trans-membrane property as delineated by Kyte-doolittle plot and (B)(ii) is a schematic to portray the positioning and transmembrane feature of the EBL S-2 portion.

3.2.3.1. Net charge contribution to form structured region

Mostly, this disulfide bonded structure imparts the ability to this structured portion to act as a hook or anchor to latch on to the membrane, as can be validated via hydropathy plots by Kyte-Doolittle method of this region and also its charge measurements via ExPasy. A comparative analysis of net charge (Table 7) between the IDR (U-2) and the last structured portion (S-2) of EBL in all Plasmodium species showed that S-2 consists of greater number of positively charged amino acids (i.e more hydrophobicity) those of which help them to construct a hydrophobic core resulting in an ordered or structured region S-2.

3.2.3.2. Hydropathy Plot

Hydropathy Plot actually supports the prediction obtained from the net charge analysis (Table 8), that is the hydropathy score of S-2 portion of EBL in all Plasmodium species

is positive (in most of the cases 3-5) (supplementary Figure 3) i.e., more hydrophobic amino acid residues; those can form a compact structured/ordered portion S-2.

It is to mention here that similar structures have been depicted in case of other species that were under study. However, those are not elucidated here as they follow the same trend. The ones with deviations are highlighted and analyzed with details. An overall representation is thus portrayed below.

3.2.4. Representative Diagram predicted from the above study

A probable structural form has thus been proposed for the entire EBL protein both structured and unstructured portions in case of P. vivax as depicted in Figure 6. The others follow similar trend. A hypothetical mechanism of their dimerization has been laid out in Figure 6 as well to understand their mechanistic aspect of function.

MAEBL, as found in P. falciparum, paralogous to erythrocyte binding proteins (EBL), essentially required for merozoite invasion of erythrocytes, a third transmembrane ligand with importance in the process of salivary gland invasion, however shows quite a strikingly different property. The usual distribution of Intrinsically Disordered Region as found in Region II (as designated in most cases) is missing in this case (Figure 7A and B). A Ramachandran Plot is depicted in Figure 7C to show the striking difference comparative to the ones corresponding to the Intrinsically Disordered Regions. The allowed and favoured Proline regions were sparse, indicating stability in the structure and lesser probability of disordered-ness. Due to the massive assembly of residues to make MAEBL makes it break down into domain versions and thereby act depending on the plan of action charted out by the protein.

Table 7: Comparison of the percentage of PEST sequence between the U-2 and S-2 regions of *Plasmodium* EBL indicates that U-2 is PEST rich i.e., PEST is an identification marker of IDP.

Plasmodium sp.	Percentage of PEST in U-2	Percentage of PEST in S-2
P. vivax	8.57	6.27
P. falciparum	8.27	7.50
P. yoelii	8.77	5.40
P. inui	7.32	5.7
P. malariae	6.55	3.40
P. knowlesi	8.70	7.20
P. reichenowi	8.40	6.60
P. cynomolgi	8.20	5.95
P. coatneyi	7.87	6.15
P. gaboni	10.07	5.40
P. fieldi	7.60	6.6
P. simiovale	7.77	6.9

Table 8: Comparisons of net charge between the U-2 and the S-2 region indicates that S-2 (net charge more positive) contains more hydrophobic amino acid residues than that of U-2 (net charge more negative).

Plasmodium sp.	Net charge in U-2	Net charge in S-2
P. vivax	-25.3	-4.1
P. falciparum	-28.8	0.7
P. yoelii	-18.6	-2.1
P. inui	-20.6	-2.2
P. malariae	-4.3	1.2
P. knowlesi	-41.3	-5.3
P. gaboni	-50.7	0.9
P. cynomolgi	-5.1	2.9
P. coatneyi	-26.9	-2.6
P. reichenowi	-41.0	-2.1
P. fieldi	-15.8	-3.9
P. simiovale	-16.3	-2.9

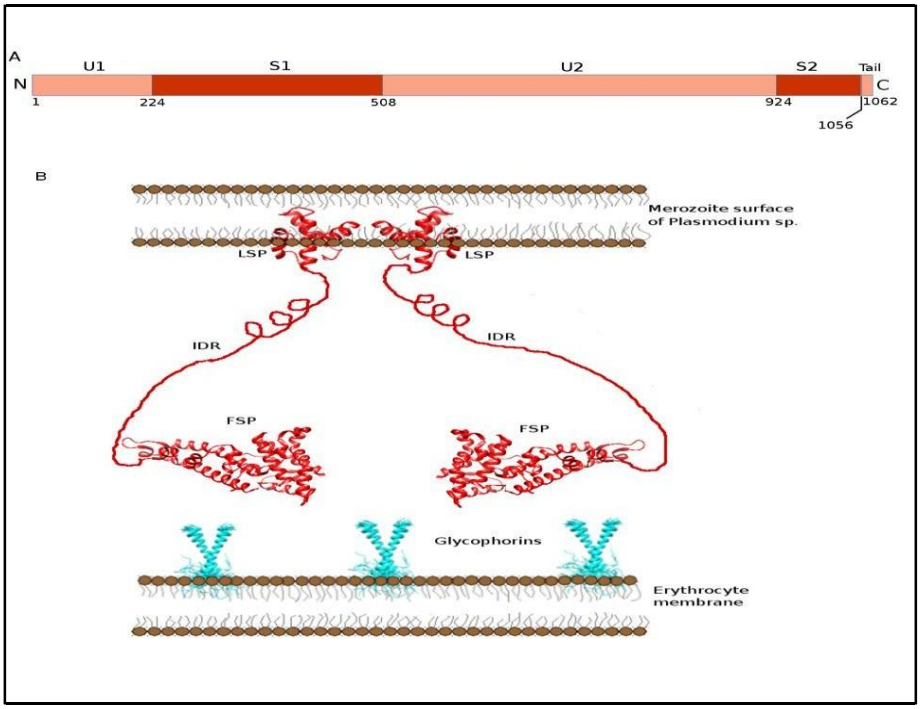

Figure 6: (A) Schematic of primary structural organisation of full length EBL of Plasmodium vivax. (B) Model showing predicted structural organization of general merozite membrane bound EBL and its interactions pattern with Glycophorin receptor located on the erythrocyte membrane. The schematic shows two membrane bound EBLs; each of their membrane embedded C-terminal S-2 or Last Structured Portion (LSP) portion is linked with the N-terminal DBD or First Structured Portion (FSP) through the flexible IDR. Dimerization may happen between DBDs upon interaction with Glycophorin, a concept adopted from a previous study (18).

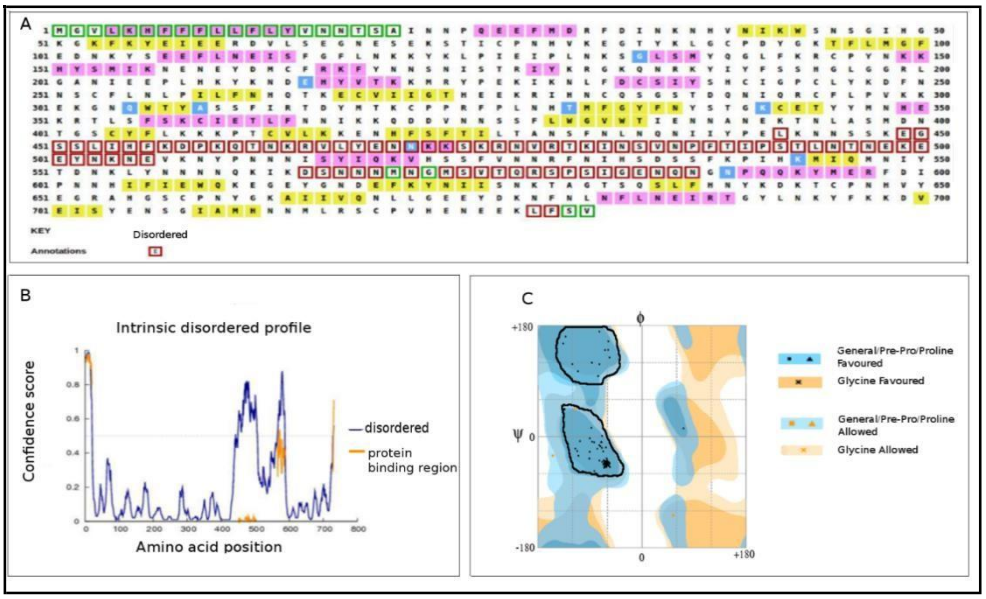

Figure 7: (A) & (B) The disordered percentage is very less in MAEBL of P. falciparum and hence predicted to have a different mode of action. (C) Ramachandran Plot depicting sparse occupancy of allowed and favoured regions of Proline residues, depicting stable structural features for MAEBL.

The unique characteristic of the Intrinsically Disordered Region and the role played by it in mediating the function of EBL made us intrigued to come up with its plausible functional implications.

Functional Implication: Dur to IDR, as validated through the structural hypothetical model, the regions (as obtained in this finding) offer extensibility, flexibility and plasticity. All these factors attribute that EBL protein, with circumstantial pressure, has the potency to evolve as per requirement to modulate its end structures to reach the EBL receptors and thereby disseminate its functions.

This opens a new vista to explore this IDR of EBL region for drug target. The structural revelation may help in creating rigidity in the flexible IDR zone thus limiting its functional mediation. Binding sites along with this finding is presented just to highlight on the idea that these pockets may serve as drug sites for creating instability to the structure- by creating stearic hindrances (for bulky molecules), for creating kinks (with proline/derivative amino acids), for disrupting its orientational/conformational flexibility by altering polarity / acidity, basicity or neutrality of the amino acid residues in that particular position of the structure.

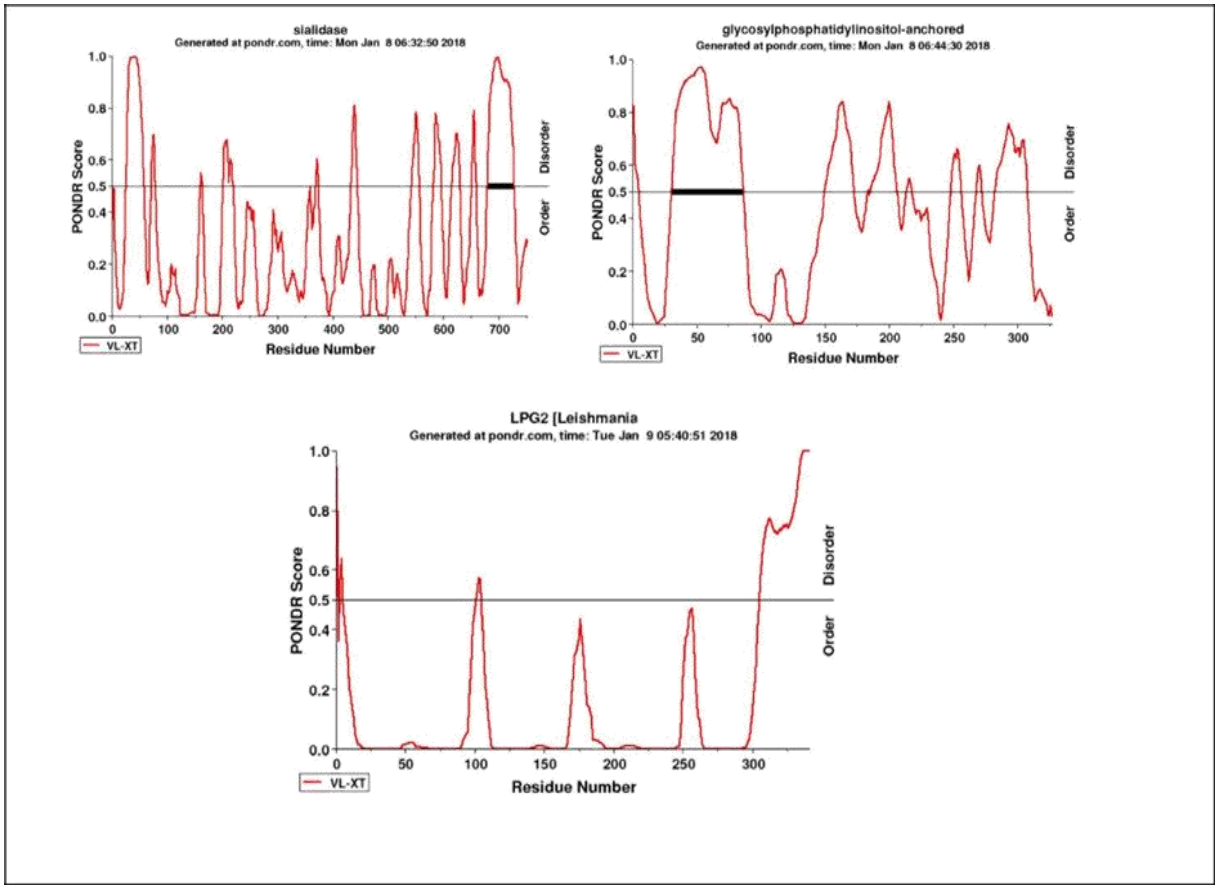

Figure 8: A: Invasion ligand Sialidase of Trypanosoma cruzi, B: glycosylphosphatidylinositol-anchored protein of Babesia divergens and C: LPG2 of Leishmania donovani- all contains certain disordered or intrinsically major portions of disordered regions (mentioned in text).

4. Discussion

The EBL protein, which plays a major role in the Plasmodium species in mediating one of the most infested diseases- Malaria, has been known to channelize its infection or disease via two routes- Erythrocyte binding ligand and Reticulocyte binding ligands. While various variants of EBL of P. falciparum were already identified and some were partially or fully characterized, the other infectious species of Plasmodium genus still remains to be fully elucidated. Through the screening and homology based bioinformatical approach, the basis of mechanism and function of the single variants were of EBL were approached for exploration, as present in most of the malarial species, identifying more similarity with P. vivax. Interestingly, they show a huge amount of Intrinsically Disordered region in their structure that is found to be quite unique to their invasion properties. MAEBL however, works a bit differently than the other ligands. Having a huge protein structure has led it to break itself down probably into domain wise form to anchor and infect in different pattern. EBL present in the examined species of Plasmodium were found to have evolutionary properties owing to the presence of a stretch of uncharacterized IDP regions whose plasticity increases its evolutionary potential. Moreover, study on its properties and probable functions along with its binding partners can actually give a far-fetched insight in correlation to how it can circumvent all the anti-malarial drugs and can even help to probe into the site to be effectively used for newer drug design. Moreover knowing the entire structure and conformations can lead to dissect further into structure-function relations of the proteins and help to understand how the adjoining structured portions help in the IDP region's integrity and plasticity simultaneously. Another similar pattern was observed in case of RBP1 and RBP2 of P. vivax and the pattern was also observed in the homologues of the proteins in P. falciparum. The mystery and purpose of the abundantly available IDRs are not yet known but we can hypothesize that it can contribute immensely to the dynamicity of the structure and thereby may have evolutionary advantages for invasion strategies. This is a new field to inspect further and identify the answers to the yet unsolved questions. For the known PDB structure of EBL of P. vivax, certain cleft and tunnel analyses were done, in presence of ligands to monitor and give an overview as to how these can actually be used as drug targets or may influence the dynamicity of the complex directly or indirectly, being associated with or as a downstream process of the IDRs. However, that needs further introspection and analyses.

Further investigations led to invasion ligands of certain other blood-borne parasites like Trypanosoma cruzi, Babesia divergens and Leishmania donovani in order to find out whether they possess the characteristic feature of being Intrinsically Disordered or not. These parasites are responsible for causing fatal infectious diseases like Chagas disease, Babesiosis and Leishmaniasis respectively (63-66). The invasive ligands studied are sialidase, BD37 & RON2 (GPi-anchored proteins), LPG2 respectively for each of these parasitic blood-borne diseases. However, the striking feature is that they also displayed certain extent of disordered-ness (Figure 8), concluding to which it can be mentioned that- the flexibility and dynamicity of the invasive ligands of Plasmodium sp. though contribute majorly for its functional properties, it is probably a characteristic feature of invasion ligands (having intrinsic disordered property of a portion of the entire protein) in order to mediate infection.

The intrinsic disordered nature of EBL thus reflects the unique feature of these proteins which enables it to be more dynamic and can transiently act as a medium to help in evolution of these ligands to mediate and spread rapid infection.

5. Acknowledgements

The authors acknowledge IISER-Kolkata for providing research facility and International Institute of Innovation and Technology for infrastructural support.

6. Author contribution

Experiments were designed by AA. Experiments were performed by AA. Experimental data were analyzed by AA. Manuscript was written by AA.

7. Conflict of interest

The author declares to have no conflicts of interest for the contents of this article.

8. Abbreviations used

Erythrocyte Binding Antigen (EBA), Erythrocyte Binding Ligand (EBL), Intrinsically Disordered Protein (IDP), Intrinsically Disordered Regions (IDR), Duffy Binding Domain (DBD).

9. Availability and implementation

No software developed. Data with the authors.

10. Supplementary information

Supplementary data are available.

11. References

1. http://www.who.int/gho/malaria/epidemic/deaths/en/
2. Chan JA, Fowkes FJ, Beeson JG. Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. Cellular and Molecular Life Sciences. 2014 Oct 1;71(19):3633-57.
3. Orlandi PA, Sim BK, Chulay JD, Haynes JD. Characterization of the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum. Molecular and biochemical parasitology. 1990 May 1;40(2):285-94.
4. Gilberger TW, Thompson JK, Triglia T, Good RT, Duraisingh MT, Cowman AF. A Novel Erythrocyte Binding Antigen-175 Paralogue fromPlasmodium falciparum Defines a New Trypsin-resistant Receptor on Human Erythrocytes. Journal of Biological Chemistry. 2003 Apr 18;278(16):14480-6.
5. Mayer DG, Mu JB, Kaneko O, Duan J, Su XZ, Miller LH. Polymorphism in the Plasmodium falciparum erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. Proceedings of the National Academy of Sciences. 2004 Feb 24;101(8):2518-23.
6. Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, Cowman AF. Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nature medicine. 2003 Jan;9(1):87-92.
7. Narum DL, Fuhrmann SR, Luu T, Sim BK. A novel Plasmodium falciparum erythrocyte binding protein-2 (EBP2/BAEBL) involved in erythrocyte receptor binding. Molecular and biochemical parasitology. 2002 Feb 1;119(2):159-68.
8. Thompson JK, Triglia T, Reed MB, Cowman AF. A novel ligand from Plasmodium falciparum that binds to a sialic acid-containing receptor on the surface of human erythrocytes. Molecular microbiology. 2001 Jul;41(1):47-58.
9. Mayer DG, Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, Mendoza LH, Miller LH. Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. Proceedings of the National Academy of Sciences. 2009 Mar 31;106(13):5348-52.
10. Jaśkiewicz E, Graczyk J, Rydzak J. Proteins involved in invasion of human red blood cells by malaria parasites. Postepy higieny i medycyny doswiadczalnej (Online). 2010 Nov;64:617-26.
11. Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. Nature. 2006 Feb;439(7077):741-4.
12. Orlandi PA, Klotz FW, Haynes JD. A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac (alpha 2-3) Gal-sequences of glycophorin A. The Journal of cell biology. 1992 Feb 15;116(4):901-9.
13. Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum. Science. 1994 Jun 24;264(5167):1941-4.
14. Sim BK, Orlandi PA, Haynes JD, Klotz FW, Carter JM, Camus D, Zegans ME, Chulay JD. Primary structure of the 175K Plasmodium falciparum erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion. The Journal of cell biology. 1990 Nov;111(5):1877-84.
15. Orlandi PA, Klotz FW, Haynes JD. A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac (alpha 2-3) Gal-sequences of glycophorin A. The Journal of cell biology. 1992 Feb 15;116(4):901-9.
16. Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, Cowman AF. Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nature medicine. 2003 Jan;9(1):87-92.
17. Lopaticki S, Maier AG, Thompson J, Wilson DW, Tham WH, Triglia T, Gout A, Speed TP, Beeson JG, Healer J, Cowman AF. Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites. Infection and immunity. 2011 Mar 1;79(3):1107-17.
18. Batchelor JD, Zahm JA, Tolia NH. Dimerization of Plasmodium vivaxDBP is induced upon receptor binding and drives recognition of DARC. Nature structural & molecular biology. 2011 Aug;18(8):908.
19. Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell. 2005 Jul 29;122(2):183-93.
20. Marti M, Baum J, Rug M, Tilley L, Cowman AF. Signal-mediated export of proteins from the malaria parasite

to the host erythrocyte. The Journal of cell biology. 2005 Nov 21;171(4):587-92.
21. Guy AJ, Irani V, MacRaild CA, Anders RF, Norton RS, Beeson JG, Richards JS, Ramsland PA. Insights into the immunological properties of intrinsically disordered malaria proteins using proteome scale predictions. PloS one. 2015;10(10).
22. https://www.rcsb.org/
23. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J. Intrinsically disordered protein. Journal of molecular graphics and modelling. 2001 Feb 1;19(1):26-59.
24. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nature reviews Molecular cell biology. 2005 Mar;6(3):197-208.
25. Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Current opinion in structural biology. 2008 Dec 1;18(6):756-64.
26. Pullen RA, Jenkins JA, Tickle IJ, Wood SP, Blundell TL. The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. Molecular and cellular biochemistry. 1975 Jul 1;8(1):5-20.
27. CARY PD, MOSS T, BRADBURY EM. High-resolution proton-magnetic-resonance studies of chromatin core particles. European journal of biochemistry. 1978 Sep;89(2):475-82.
28. Forman-Kay JD, Mittag T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure. 2013 Sep 3;21(9):1492-9.
29. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK. D2P2: database of disordered protein predictions. Nucleic acids research. 2012 Nov 29;41(D1):D508-16.
30. Jakob U, Kriwacki R, Uversky VN. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. Chemical reviews. 2014 Jul 9;114(13):6779-805.
31. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: structure, function, and bioinformatics. 2000 Nov 15;41(3):415-27.
32. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins: Structure, Function, and Bioinformatics. 2001 Jan 1;42(1):38-48.
33. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry. 2014 Jun 2;83:553-84.
34. Adzhubei AA, Sternberg MJ. Left-handed polyproline II helices commonly occur in globular proteins. Journal of molecular biology. 1993 Jan 20;229(2):472-93.
35. Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. Nature reviews Molecular cell biology. 2006 Jul;7(7):473-83.
36. Arif M, Senapati P, Shandilya J, Kundu TK. Protein lysine acetylation in cellular function and its role in cancer manifestation. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2010 Oct 1;1799(10-12):702-16.
37. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. Nature structural & molecular biology. 2010 Jun;17(6):666-72.
38. Bicker KL, Thompson PR. The protein arginine deiminases: Structure, function, inhibition, and disease. Biopolymers. 2013 Feb;99(2):155-63.
39. Wright PE, Dyson HJ. Linking folding and binding. Current opinion in structural biology. 2009 Feb 1;19(1):31-8.
40. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends in biochemical sciences. 2008 Jan 1;33(1):2-8.
41. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure. 2010 Mar 14;18(4):494-506.
42. Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. Nature reviews Molecular cell biology. 2004 Nov;5(11):908-19.
43. Young JC, Agashe VR, Siegers K, Hartl FU. Pathways of chaperone-mediated protein folding in the cytosol. Nature reviews Molecular cell biology. 2004 Oct;5(10):781-91.
44. Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. The FASEB Journal. 2004 Aug;18(11):1169-75.
45. Ivanyi-Nagy R, Davidovic L, Khandjian EW, Darlix JL. Disordered RNA chaperone proteins: from functions to disease. Cellular and Molecular Life Sciences CMLS. 2005 Jul 1;62(13):1409-17.
46. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature. 2007 Jun;447(7147):1021-5.
47. Wright PE, Dyson HJ. Linking folding and binding. Current opinion in structural biology. 2009 Feb 1;19(1):31-8.
48. Wu H. Higher-order assemblies in a new paradigm of signal transduction. Cell. 2013 Apr 11;153(2):287-92.
49. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science. 2008 Nov 28;322(5906):1365-8.
50. Dobson CM. Protein folding and misfolding. Nature. 2003 Dec;426(6968):884-90.
51. https://www.ncbi.nlm.nih.gov/
52. https://swissmodel.expasy.org/
53. http://www.ebi.ac.uk/Tools/msa/clustalw2/
54. http://bioinf.cs.ucl.ac.uk/psipred/
55. http://www.pondr.com/
56. ProtCalC.sourceforge.net
57. http://web.expasy.org/protparam/
58. http://www.uniprot.org/
59. http://bioware.ucd.ie/~compass/biowareweb/Server_pages/slimpred.php
60. http://web.expasy.org/protscale/
61. http://anchor.enzim.hu/
62. http://mordred.bioc.cam.ac.uk/~rapper/rampage.php

63. Schenkman S, Eichinger D. Trypanosoma cruzi trans-sialidase and cell invasion. Parasitology Today. 1993 Jun 1;9(6):218-22.
64. Lobo CA. Babesia divergens and Plasmodium falciparum use common receptors, glycophorins A and B, to invade the human red blood cell. Infection and immunity. 2005 Jan 1;73(1):649-51.
65. Lobo CA, Cursino-Santos JR, Rodriguez M, Ord R, Singh M. Receptor–ligand interactions underlying RBC invasion in Babesia. ISBT Science Series. 2015 Apr;10(S1):173-80.
66. Späth GF, Garraway LA, Turco SJ, Beverley SM. The role (s) of lipophosphoglycan (LPG) in the establishment of Leishmania major infections in mammalian hosts. Proceedings of the National Academy of Sciences. 2003 Aug 5;100(16):9536-41.