In-Silico design of peptide inhibitors against PfEMP1 Cerebral Malaria

AUTHOR
Anantha S Rao
20181044

III BS-MS student, IISER Pune

SUPERVISOR
Prof. Sanjeev Galande

Division of Biology, IISER Pune

September 2020 - December 2020

Contents

1 Abstract

Malaria, the mosquito borne disease, is one of the toughest global health crises we face today. In 2018, it was responsible for 228 million infections and 405,000 deaths worldwide, with children alone accounting for 67% of the total deaths. [2]As of today, Artemisinin Combination Therapy (ACT) is the most potent anti-malarial drug, but their is growing evidence for malarial parasites attaining Artemisinin resistance worldwide, [5] potentially making malaria untreatable in the years to come.

We computationally designed a library of peptide drugs against *Plasmodium falciparum*, particularly targeting the interaction of the DBLb domain of erythrocyte membrane protein, PF110521 with the Human Endothelial receptor, Intercellular Adhesion Modelcule-1 (ICAM-1). The designed peptide inhibitors were grafted in loop 6 of Cyclotide kalata B1, a circular plant protein known for its thermal stability, anti-peptidase activity and utility as an efficient drug scaffold. The interactions of peptide inhibitors were scored using an empirical force field and their binding stability was accessed using Molecular Dynamics Simulations.

This project is a part of a larger iGEM project (IISER Pune, 2020) [1] that aims to create a library of inhibitory peptide molecules, which target multiple, host-pathogen interactions in the blood stage of *Plasmodium falciparum* malaria, effectively preventing the parasites from infecting human red blood cells.

2 Introduction

PfEMP1 (Plasmodium falciparum Erythrocyte Membrane Protein 1) are a class of membrane proteins exhibited by P.falciparum Infected Erythrocytes (IE) and coded by ~ 60 var genes. The sequence of var genes expressed during an infection cycle of P.falciparum are still unknown, and are said to have evolved to avoid protective immune responses from the host.

The immunopathology of P.falciparum malaria has been linked to the cyto-adhesion of IE to specific host receptors. This IE adhesion is called 'Sequestration' when the IEs stick to tissue-bound receptors, 'Rosetting' when they stick to uninfected erythrocytes, and 'Clumping' when the IEs stick to each other. The primary function of the PfEMP1 proteins is to mediate this cyto-adhesion and assist in the proliferation of IEs in the vasculature.[7]

Lennartz et al found that a region of A-type PfEMP1 is associated with an increased risk of developing symptoms of Cerebral Malaria, an advanced stage of malaria characterised by the IEs crossing the blood-brain barrier and accumulating in the brain. The authors developed the crystallographic structures of the DBLb Domain of PfEMP1 interacting with the human ICAM-1 endothelial receptor (PDB ID : 5MZA) and found certain specific motifs responsible for Human ICAM-1 binding, that are conserved across all var genes. The ICAM-1 binding was found to be mediated mainly by a subset of Duffy binding-like (DBL) domains called the DBL-b domains in PfEMP1A. [8] We used this interaction to design peptide inhibitors against the DBL-b Domain of PfEMP1A protein.

The binding of A-type PfEMP1 with ICAM-1 is possible due to 3 main binding regions of ICAM-1:

1. Region 1: A Hydrophobic-Hydrophobic interaction
2. Region 2: A motif rich in Glycine-Proline that allows sharp backbone bends and formation of hydrogen bonds with ICAM-1. Representing the motif with a single-letter-aminoacid code and 'x' for any amino acid (Prosite notation), we have :

$$\mathrm{I[V/L]x_3N[E]GG[P/A]xYx_{27}GPPx_3H}$$

3. Region 3: An intrinsically disordered region of PfEMP1A responsible for variable binding.

We designed peptide inhibitors against the motifs that form hydrogen bonds with ICAM-1, namely region 2 since this is the strongest of the three interactions. We hypothesize that inhibiting this interaction with a drug that competitively binds to P.falciparum protein will prevent the binding of infected erythrocytes with Human-receptors and aid in the prevention of Cerebral Malaria.

3 Materials and Methods

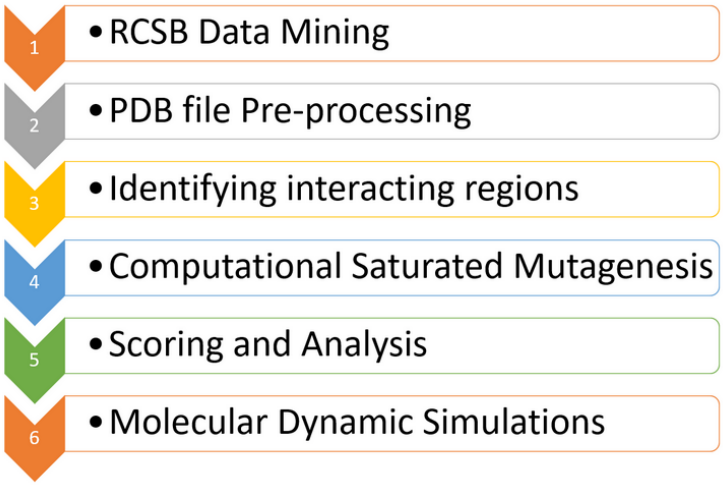


Figure 1: Summary of the Workflow followed

3.1 RCSB Data-Mining and Preprocessing of PDB files

Template-based drug design primarily requires an experimentally solved structure of the epitope protein against which drugs are to be designed. Using the deposited PDB files, we can further look at designing small molecule and peptide drugs using knowledge-based methods. For this purpose, we started with an advanced search on the RCSB database and looked at all structures containing a *falciparum* protein and a human protein. We wrote a static-web-parser to read the RCSB database for matching items and obtained 444 hits. Since the preliminary search contained many structures with artificial antibodies, we further restricted our search to wild-type Human and *falciparum* proteins and ended up with 5-6 strucutres of different crucial interactions in the blood stage of falciparum malaria. Out of them, I worked on the PDB ID : 5MZA.

The strategy we used to inhibit the formation of the Host-Parasite protein complex was to design a peptide that would competitively bind to the Parasite protein, preventing its

binding to the host (Human) protein. Therefore, we identified motifs on the Human protein that are closest to and bind strongly to the parasite protein. We isolated these motifs along with the parasite protein into different structures and considered each one of them as a potential inhibitory model. By this method, we generated multiple models each consisting of the parasite protein and a peptide sequence (6-20 amino acids in length) from the Human protein.

3.2 Computational Saturated Mutagenesis and Scoring

Saturation mutagenesis is a protein-engineering technique in which single/multiple aminoacid(s) are substituted with all possible amino acids at a particular position. This is used widely to improve upon certain properties of proteins such as catalytic activity, enantioselectivity, thermostability, and binding affinity. [9]

We performed Saturation Mutagenesis on the peptide side-chain of each inhibitory model to obtain mutated sequences. This was done using a Python 2.7 script that acts as a wrapper function over UCSF Chimera [10] to mutate a specific residue into a different side-chain conformations or into a different amino acid for each position of the original sequence. This was repeated for all the inhibitory models.

To compare the binding energies of all the mutant peptides with the parasite protein, we scored all the the models obtained after saturation mutagenesis using FoldX [12], an empirical force field developed for the rapid evaluation of stability, folding, and dynamics of proteins. We worked with FoldX because it had been validated on multiple protein-peptide complexes, to estimate the change of the protein–peptide-binding energy with respect to the energy of an experimentally resolved complex and also because the use of FoldX in protein design has been extensively reported. [13]

3.3 Molecular Dynamics Simulations

Molecular dynamics is a coarse-grained computer simulation method used to study atoms and molecules. The technique is based on the generation of atomic trajectories of a system over time by numerical integration of Newton's laws of motion for a specific inter-atomic potential defined along with appropriate initial and boundary conditions. [11]

Molecular Dynamics Simulations were performed to understand the interaction between the peptide and the *falciparum* protein. If the binding affinity is high, then peptide acts as an efficient inhibitor and can slow down the pathogenesis of *Plasmodium falciparum*. Molecular dynamics was also used to understand the stability of the grafted inhibitor in a synthetic kalataB1 cyclotide. The GROMACS 2019.1 package. [3] with the AMBER99SB-ILDN force field was used to perform MD simulations. The AMBER99SB-ILDN force field was used because it had a better accord with experimental NMR relaxation data of test protein systems and also was used for the MD Simulation of a similar study of peptide-peptide interaction. [6]. All MD Simulations were performed on the PARAMBRAHMA facility under the National Supercomputing Mission, Government of India at IISER Pune.

3.3.1 Analysis of Molecular Dynamics Data

From the large volume of MD simulation data, we wanted to study two particular aspects in greater detail:

1. **Centroid Profile of the Inhibitor**
   Here, we want to study how the distance between the centroids of the protein and peptide evolve over time.

   (a) If the distance between the centroids is increasing, then the peptide moves away from the protein and is not binding effectively.

   (b) If the distance between the centroid of the two entities remains constant, then over the duration of the simulation, the peptide binds to the protein and can act as a good inhibitor.

(c) If the distance decreases, then the peptide moves closer to the Protein core. (However, this is assumed to be unlikely). To perform these calculations, I wrote the script PDB-centroid-analyser.py that computed the centroid profile for a given inhibitor.

2. **Hydrogen Bond Profile of the Inhibitor**

   To study effective binding, we analysed the number of inter-molecular hydrogen bonds between the Protein and the peptide. This required calculating the hydrogen bond profile for each snapshot of the MD simulation and to determine which residues and specifically which atom in these residues of the peptide inhibitor form hydrogen bonds with the protein. We also wanted to determine the relative abundance of hydrogen bonds formed over the entire simulation period. To perform these calculations I wrote the script PDB-Hbond-analyser.py

3.4 Grafting of Inhibitors into Cyclotides

The ultimate aim of the project was to graft the designed inhibitors into loop 6 of Cyclotide kalata-B1. This required building the 3D structure of the grafted cyclotide drug and computationally assessing its stability. The 3D structure was modelled using MODELLER [4] which by default yields a non-circularised peptide. End-to-end circularisation of the backbone was achieved by using a custom script based on a special patch.

Equilibrium Molecular Dynamics simulations were performed to study the stability of the grafted cyclotide. The simulation was performed for 100ns using GROMACS with the same parameters as the protein-peptide MD simulations mentioned earlier. The Atomic Root Mean Square Deviations were calculated with respect to the protein backbone which revealed the evolution of average inter-atomic-distance with time.

4 Results

4.1 Preprocessing

The raw PDB file obtained from the RCSB database contained some missing regions and residues, namely residues no: 1027-1037, 1126-1156, 1188-1193 of Chain A. These missing regions and loops were modelled based on the principles of homology modeling. Inositol 6-phosphate and other non-standard residues that were used for crystallization were removed to aid in further analysis. The complete structure obtained is shown in Fig:2a
Five motifs of the human protein (Chain B) that were closest to the parasite protein were identified (motif + Protein) and used as models for further analysis.
The five models were:
1. Residues 10-18 : ILPRGGSVL (Fig:2b)
2. Residues 38-45 : PKKELLLP (Fig:2c)
3. Residues 49-58 : RKVYELSNVQ (Fig:2d)
4. Residues 83-88 : YWTPER (Fig:2e)
5. Residues 168-173 : QGLELF (Fig:2f)

4.2 Computational Saturated Mutagenesis and Scoring

A notation to identify all the peptides with a specific alpha-numeric code was introduced for easy reference, and data handling.
Example:

1. 1_2_A21C refers to the second model of the first interaction where residue 21 Alanine (A) was mutated to Cysteine (C)

2. 1_1_H refers to the first model of the first interaction, where a hybrid peptide was obtained. A hybrid peptide is one where all amino acids are substituted with the ones that yield maximum score.

Computational Saturation Mutagenesis was performed on all the five peptide models obtained in the previous section. After repairing clashes in all models to satisfy spatial constraints

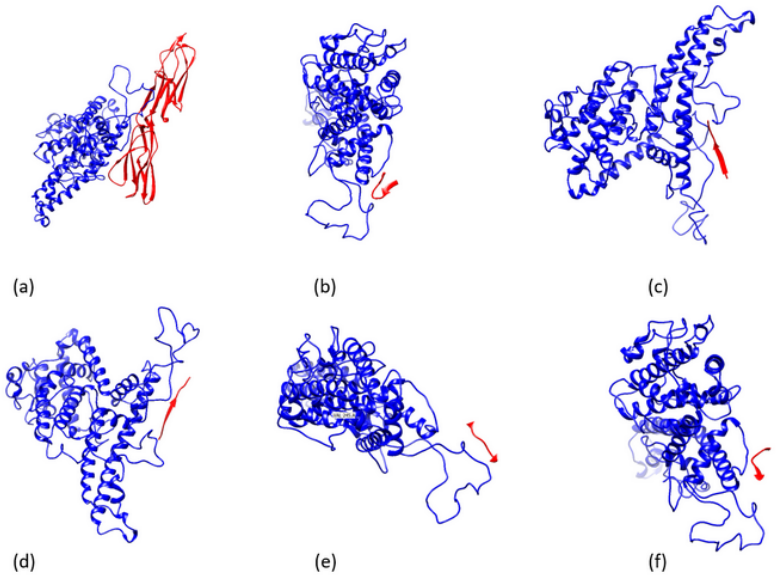


Figure 2: Inhibitory Models for the DBLb domain of PfEMP1A
Blue : Parasite Protein

with the Foldx empirical force field [12], all the unique 680 models were scored and sorted based on Interaction energy values. Mutants of Model 1 (residues 10-18) were found to have the least interaction energy and highest stability. The scatter plots of the interaction energy (Fig3a) and the stability of the peptide against each mutation illustrate the fact that inhibitory models of Model 1 are most stable. This was inline with the experimental result from Lennartz et al and was in a way, a computation based verification of the same [8]. Now on-wards, we decided to focus on this peptide motif for future analysis.

The heat map of interaction energy (Fig4) scores of the mutants from the first model are shown below. The X-axis gives the position of mutation and the Y-axis shows the corresponding amino acid. Mutation of residue 16 Ile (Isoleucine) and Phe (Phenylalanine) causes the highest decrease in Interaction energy between the Parasite protein and the peptide. Furthermore mutating residue 16 (wild type - Serine) influences the stability of binding the highest while mutation of residue 15 (wild type - Glycine) tends to destabilize the interaction.

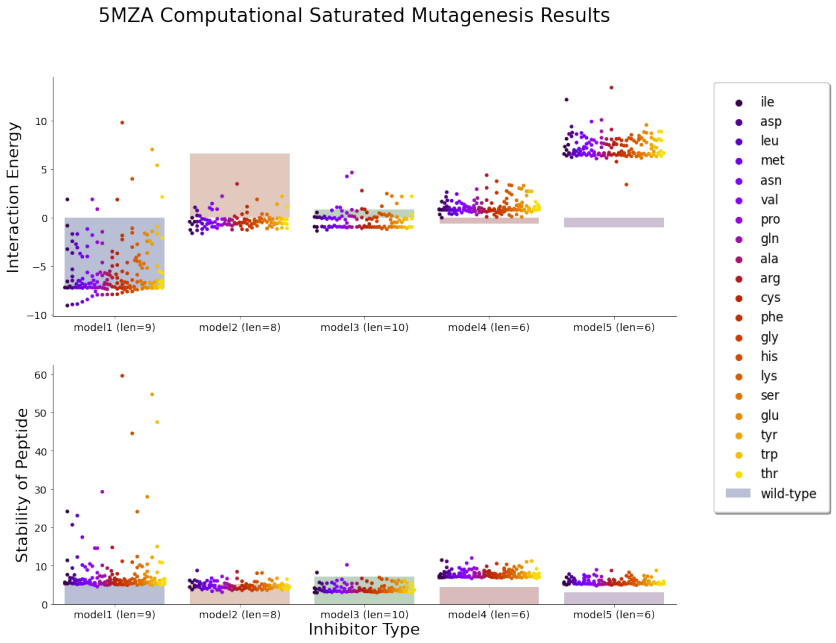


Figure 3: Results of Computational Saturated Mutagenesis

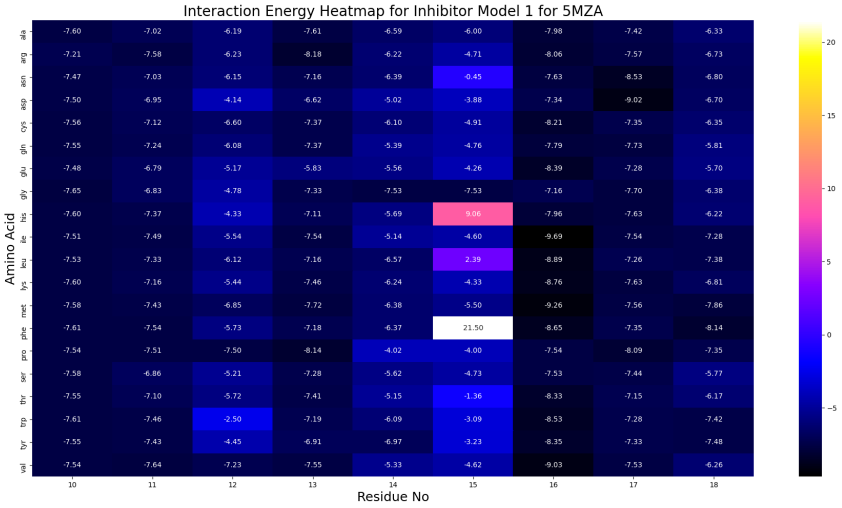


Figure 4: Results of Computational Saturated Mutagenesis for Model 1

4.3 Molecular Dynamics Analysis

Triplet Molecular dynamics simulations were performed for the top two models with the lowest interaction scores : 1_1_S16I and 1_1_H, to determine how the peptide moved relative to the protein and if it was stable, which atoms contributed to the overall stability.

4.3.1 Inhibitor 1 : 1_1_S16I

The inhibitor considered for Molecular dynamics is obtained by mutating the 16^{th} residue of the wild type human protein from Serine (S) to Isoleucine (I) in the 1st Inhibitor model. Overall, the trajectory of the peptide over 100ns remained stable around 31$\mathring{A}$ from the core of the P. falciparum protein. However, in the second run it was noticed that when the peptide was pushed towards the Protein core, it rebounded and started to move away.

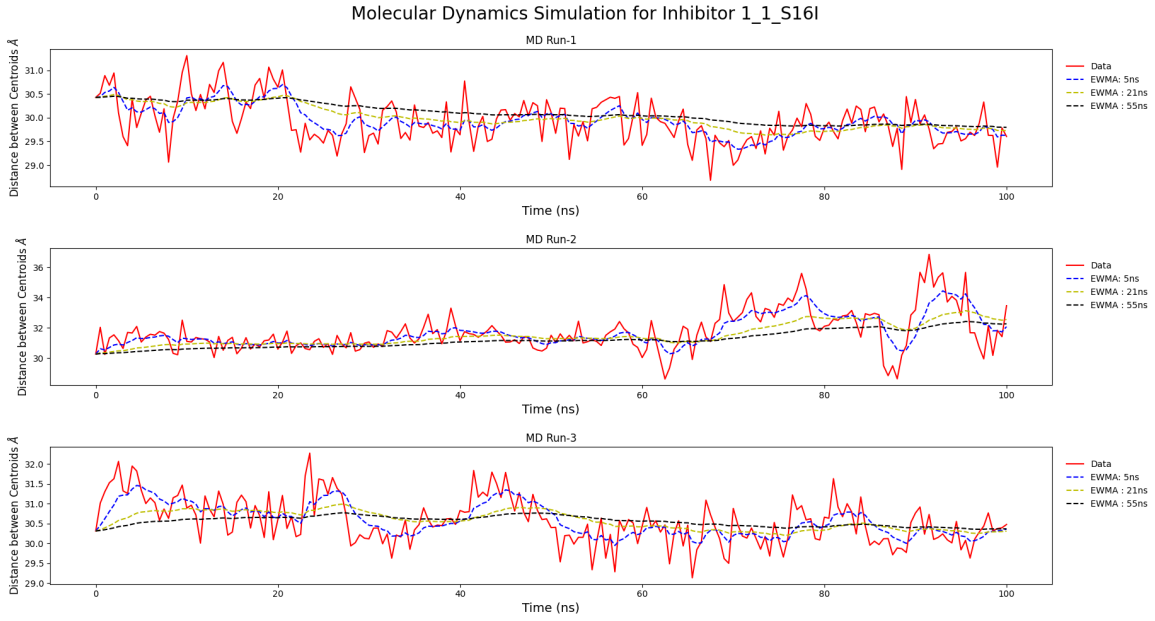


Figure 5: Evolution of distance between the peptide and the center of the protein

Exponentially Weighted Moving Average (EWMA) was used to understand the average behavior. In Fig 5 EWMAs suggest that the peptide initially bounces away from the protein

and then maintains a constant distance from the core, indicative of strong binding. The mean distance of the peptide from the protein was found to be around 31 Å and the standard deviation varied for the three runs with the second run having the highest among the three. (0.466, 9.78, 0.578) Å

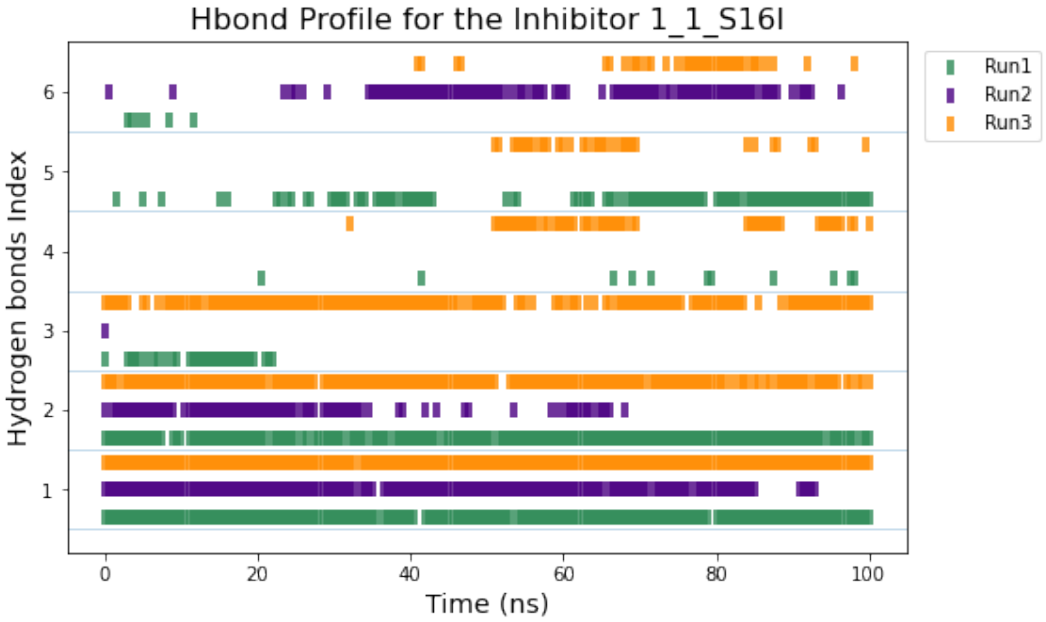


Figure 6: Conservation of Hydrogen-bonds between the peptide and the Protein

Index no	Hbond Acceptor-Donor Atom	MD run 1 (%)	MD run 2 (%)	MD run 3 (%)
1	ILE:16.B.O - ASN:356.A.ND2	97.51	84.55	99.50
2	ASN:356.A.OD1 - ILE:16.B.N	93.53	42.28	92.39
3	ARG:13.B.O - THR:394.A.OG1	15.94	<5.00	82.58
4	THR:363.A.OD1 - GLY:10.B.N	5.47	<5.00	26.86
5	GLY:10.B.O - ASN:363.A.N	52.23	<5.00	16.41
6	PHE:18.B.OC2 - LYS:349.A.NZ	<5.00	8.47	18.41

Table 1: Conservation of Hydrogen Bonds for 1_1_S16I

The Hydrogen Bond Profile (Fig 6 and Table 1) suggest that the number of inter-molecular

hydrogen bonds between the parasite protein and the peptide changed over time with up-to 3 hydrogen bonds retained for $> 80\%$ of the interaction-time. The two hydrogen bonds formed by the O-atom of the mutated Isoleucine residue at position 16 of the peptide had the strongest retention over time, thus reinstating the initial hypothesis of efficient binding due to higher number of hydrogen bonds.

4.3.2 Inhibitor 2 : 1_1_H

Model 1_1_H refers to the Hybrid inhibitor of 5MZA. The Hybrid peptide was created by mutating each amino acid in the wildtype to the one that yielded least scores on scoring. It was the model with the least Interaction energy, thereby suggesting highest stability.

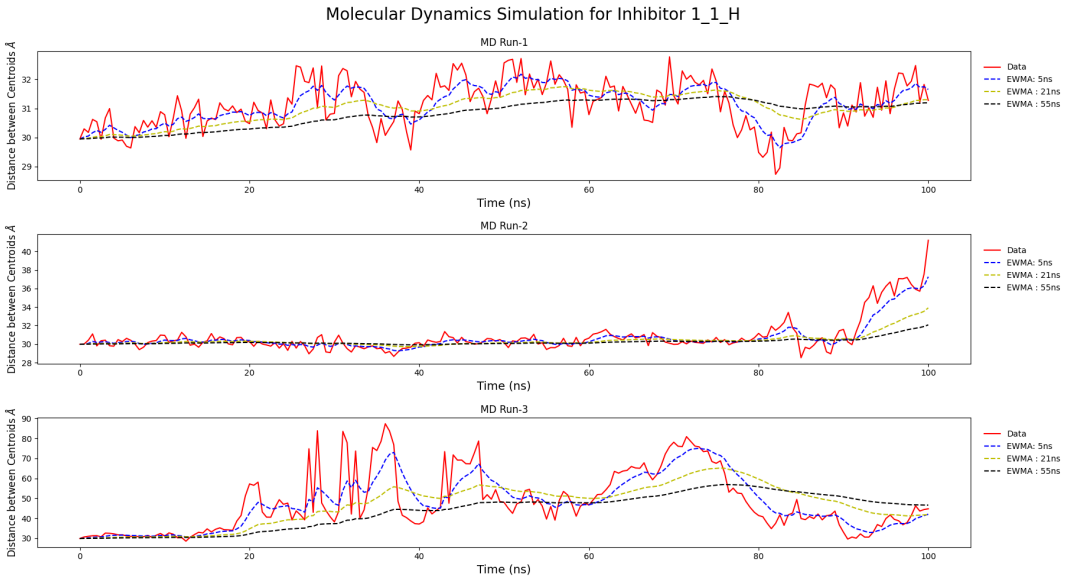


Figure 7: Evolution of distance between the peptide and the center of the protein

The distance between the centroids of the protein and peptide increased in two runs while it remained constant in another run. The peptide initially tended to move away from the protein but the distance stabilized over time. However, in runs two and three, it was found that the peptide moved too quickly and then failed to stabilize over time. The mean

distance was found to be 31.14, 32.94, 48.07 $\mathring{A}$ with standard deviations 0.811, 11.72, 14.97 $\mathring{A}$ respectively for run 1, 2 and 3.

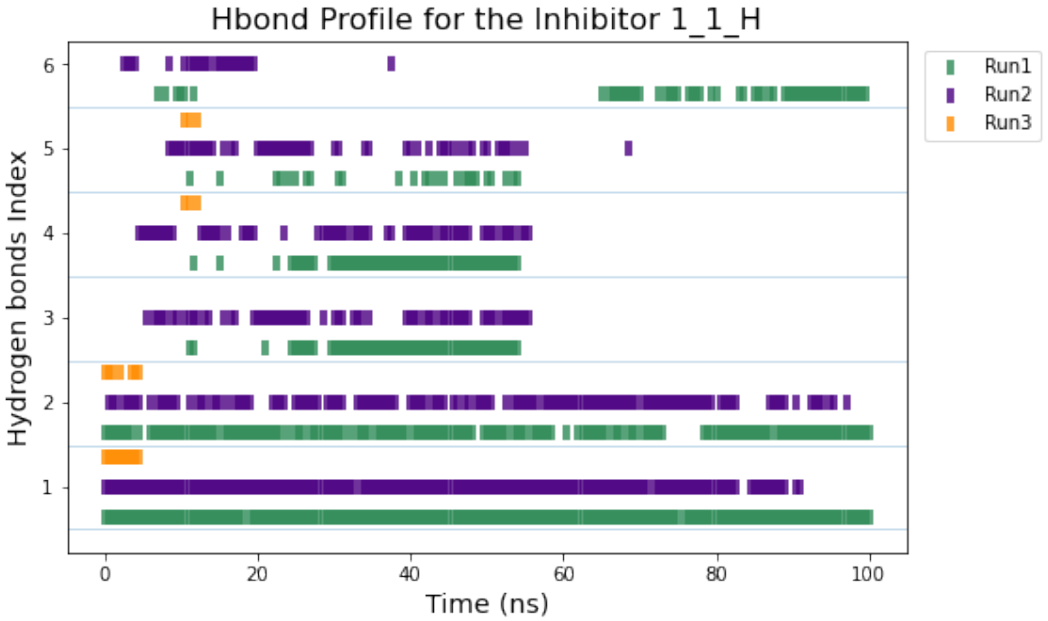
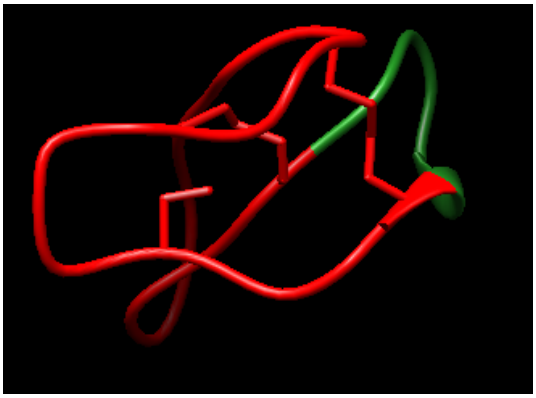


Figure 8: Conservation of Hydrogen-bonds between the peptide and the Protein

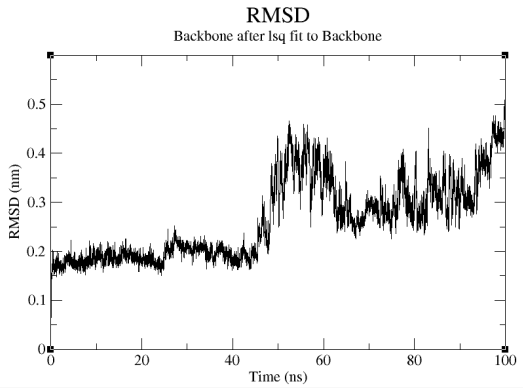
Index no	Hbond Acceptor-Donor Atom	MD run 1 (%)	MD run 2 (%)	MD run 3 (%)
1	ILE:16.B.O - ASN:356.A.ND2	99.05	87.56	<5.00
2	ASN:356.A.OD1 - ILE:16.B.N	83.08	71.14	<5.00
3	ASP:393.A.OD2 - ARG:13.B.NE2	29.85	32.83	<5.00
4	ASP:393.A.OD1 - ARG:13.B.NH2	29.85	31.84	<5.00
5	ASP:393.A.OD2 - ARG:13.B.NH2	13.45	26.86	<5.00
6	ASP:17.B.OD2 -TYR:362.A.OH	27.86	11.94	<5.00

Table 2: Conservation of Hydrogen Bonds for 1_1_H

The Hydrogen Bond profile for the three runs (Fig 8 and Table 2), confirmed that the peptide effectively flew away in the 3rd MD run. However, we obtained conclusive evidence from Run1 and Run2 that Isoleucine at residue 16, formed two new hydrogen bonds with the



(a) 1_1_S16I grafted into loop-6 of kalata-B1



(b) RMSD Plot for 1_1_S16I grafted into loop-6 of kalata-B1

Figure 9: Insilico Grafting Results

protein and was conserved for almost $\sim 93\%$ of the simulation period.

4.3.3 Grafting Results

Based on the promising results of 1_1_S16I, the peptide sequence was grafted in loop 6 of Cyclotide kalata B1 while a Strep tag (WSHPQFEK) was grafted in loop 3 of kalata B1 to aid in the purification of the cyclotide. The 3D structure was modelled as described in the methods section. Fig 9a shows the insert (1_1_S16I) depicted in green while the cyclotide CCK-backbone along with the Strep tag is depicted in red.

Using the MD data, RMSD calculation (Fig 9b) was performed with respect to the minimized, equilibrated system of this grafted cyclotide. In the trajectory movie, the peptide remained stable and in its native folded state for a long time before starting to unfold. We believe that unfolding and other unwanted conformations could be minimised by using better modelling techniques.

5 Conclusion and Discussions

1. Overall, in this study we found multiple peptide inhibitors against the DBL-b domain of PfEMP1A protein that are responsible for binding to the Human ICAM-1 receptor and assisting in the transfer of Infected erythrocytes across the blood-brain barrier. We used diverse computational tools and techniques to design, score and test peptide inhibitors. We found that the peptide 1_1_S16I (ILPRGGIVL) and 1_1_H (GVPRGGIDF) exhibited the strongest and most stable interaction with the protein.

2. From Molecular Dynamics (Fig5), we found that the average distance of the peptide 1_1_S16 remained stable around 31 $\mathring{A}$ from the protein, while Fig6 confirmed that multiple hydrogen bonds were formed and retained during the entire time of the simulation. These results are concurrent with those of Lennartz et al [8]. Results for the Hybrid peptide indicated that the peptide did not bind effectively to the Protein. (Fig 7)

3. The grafting results from Fig 9b showed large deviations in RMSD values. Due to lack of time, we could not effectively search for alternate methods to circularize proteins and had to use an old patch written for MODELLER. We believe that a better modelling approach can be employed for the same.

4. An important aspect to consider while grafting peptides in the cyclotide is the relative sizes of the former and the later. Cyclotide Kalata B1 is small-sized peptide with 29 aminoacids and our insert is nine amino-acids in length. A larger graft will tend to destabilize the CCK-backbone while a smaller graft might be non-specific. A future direction to test and optimize the stability would be to vary the length of the graft by removing aminoacids from the ends and searching for an optimum.

5. To confirm and validate these findings, we directed the wet-lab team to build primers and biobricks for the Inhibitor 1_1_S16I and test for immunogenecity of the peptide drug and its binding with the DBL-b domain of PfEMP1A.

6 Contribution and Acknowledgements

This project was done as part of the 2020 iGEM IISER Pune project "Anopheles: The Half-Blood Princess" (https://2020.igem.org/Team:IISER-Pune-India). The idea of designing peptide inhibitors against Malaria and using a cyclotide drug scaffold was conceived by the team and it was built upon and polished by the whole team. I would like to thank my fellow team members and mentors for helping me with the brainstorming, troubleshooting errors, writing and maintaining good code practises. The code for all sections were written by me and are available on this Github Repository

I would like to thank my supervisor, Dr. Sanjeev Galande for his continuous support and guidance. I would like to thank my mentors Mrs. Tejashree Kanitkar for providing valuable guidance and helping me with all the computational techniques and Mr. Krishnendu Roy, for his important insights, advice on the implementation of wet-lab experiments and scientific report-writing. I would like to thank my colleague Antony Kiran for helping me write the MODELLER script for developing the final circularised cyclotide graft.

I would like to thank the institute and the entire Department of Biology for the support I have been offered. All the code, scripts, results and plots are publicly available on the Github Repository : `https://github.com/Anantha-Rao12/Peptides-against-Cerebral-Malaria`

7 References

References

[1] Team:iiser-pune-india-2020.igem.org. (2020, october 27). igem iiser pune. https://2020.igem.org/team:iiser-pune-india.

[2] The "world malaria report 2019" at a glance https://www.who.int/news-room/feature-stories/detail/w

[3] Herman JC Berendsen, David van der Spoel, and Rudi van Drunen. Gromacs: a message-passing parallel molecular dynamics implementation. *Computer physics communications*, 91(1-3):43–56, 1995.

[4] Narayanan Eswar, Ben Webb, Marc A Marti-Renom, MS Madhusudhan, David Eramian, Min-yi Shen, Ursula Pieper, and Andrej Sali. Comparative protein structure modeling using modeller. *Current protocols in bioinformatics*, 15(1):5–6, 2006.

[5] Rick M Fairhurst and Arjen M Dondorp. Artemisinin-resistant plasmodium falciparum malaria. *Emerging Infections 10*, pages 409–429, 2016.

[6] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[7] Anja Ramstedt Jensen, Yvonne Adams, and Lars Hviid. Cerebral plasmodium falciparum malaria: The role of pfemp1 in its pathogenesis and immunity, and pfemp1-based vaccines to prevent it. *Immunological Reviews*, 293(1):230–252, 2020.

[8] Frank Lennartz, Cameron Smith, Alister G Craig, and Matthew K Higgins. Structural insights into diverse modes of icam-1 binding by plasmodium falciparum-infected erythrocytes. *Proceedings of the National Academy of Sciences*, 116(40):20124–20134, 2019.

[9] Yuval Nov. When second best is good enough: another probabilistic look at saturation mutagenesis. *Applied and environmental microbiology*, 78(1):258–262, 2012.

[10] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[11] Dennis C Rapaport. *The art of molecular dynamics simulation*. Cambridge university press, 2004.

[12] Joost Schymkowitz, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. The foldx web server: an online force field. *Nucleic acids research*, 33(suppl_2):W382–W388, 2005.

[13] Peng Xiong, Chengxin Zhang, Wei Zheng, and Yang Zhang. Bindprofx: assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. *Journal of molecular biology*, 429(3):426–434, 2017.