iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54370

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Investigating Evolutionary Pressures on Protein Sequence: Applications to Drug Resistance in Flatworm Parasites

Urvashi Verma
MSc, Aberystwyth University, Wales

Abstract: ***The aim of This report is that it focuses on a particular flatworm called S mansoni and the disease it causes in humans and cattle, called schistosomiasis. This disease causes a million deaths annually across the globe. The drug for this disease is praziquantel, which is quite effective. Though lately, S mansoni is showing resistance towards this drug. In particular, The aim of this project is to find the regions showing evolutionary pressure in its protein structure and, hence, know how the protein structure mutates. The methods used in this project, amongst others, are SNAP, rasmol, clustal omega mainly. They are supported by using fundamentals from PDB, Biomart, blast. These methods are very helpful to biologists or people with less programming experience. This report will explore each method in detail before using them, and the steps needed to properly execute them are discussed as well. Along with the methods, this will use raw data in the form of protein sequence/ FASTA sequence from biomart. For the results, 12 different types of flatworms are chosen but narrowed down to 3-4, which closely resemble the structure of S mansoni. These flatworms are combined into a multiple sequence alignment. They are later compared using SNAP tool. Using the data from SNAP, the dn/ds ratio for each codon number is computed and the type of selection is determined. {dn/ds ratio and selection type are defined in further sections} For discussion and conclusion, the selection type, if positive, shows evolutionary pressure and the protein structure mutates and discusses its inference. It is shown visually through RASMOL tool. This tool helps in locating the binding sites where mutation takes place, so, focusing on those will help future researchers in finding a universal vaccine for this and some other viruses as well. The main finding of this experiment is the detection of regions of positive selection in the protein structure of the flatworm, using dn/ds. Also, the amino acid residues are explored as well. As with every experiment, this report covers the limitations of the process and the future works being carried on. The referencing style is APA Harvard, and the appendix section covers all the raw data used.***

I. INTRODUCTION

flatworms are a type of parasites invertebrates which are mostly bilaterian, unsegmented and soft bodied. Though unlike other bilaterians, they have no circulatory and respiratory organs. This also results in their flattened shape as it allows oxygen and nutrients to pass through their bodies. Nearly 80% of the flatworms are parasitic, they live on or in something to secure nourishment from them. It has 4 classes – flukes, tapeworms, turbellaria and monogenean. Its full name is *Schistosoma mansoni.* It belongs to the blood fluke's category and type of human parasite which is water borne. It causes a major disease in humans called intestinal schistosomiasis. It is caused in the urinary tract and the intestines. Mainly symptoms include blood in urine, diarrhoea, abdominal pain. Having this disease for a long time causes kidney damage, liver damage, infertility. This disease is caused by fresh water infected by parasites. It is treated with praziquantel – a medicine used to treat parasitic infections in mammals. Neglected tropical diseases are tropical infections mainly found in Africa, Asia, Americas. They affect
around 1.5 billion people globally and result in many deaths. For this report, the focus is on S. Mantoni, and the disease caused by it - intestinal schistosomiasis, which is also classified as a neglected tropical disease. To understand the reason behind this particular parasite affecting the human body, it will analyse and compare the protein sequence/ amino acid sequence of various flatworm species. Proteins are formed by amino acids when then undergo a reaction called condensation reaction. Such reactions are important to combine 2 single molecules to form one big molecule (generally through the loss of one water molecule). So, when amino acids undergo this reaction, they lose one water molecule per reaction to attached themselves to one another in peptide bond. It is a covalent bond linking 2 alpha amino acids, each of the AA are from Carbon and nitrogen, respectively. Also known as selective pressure or natural selection, any cause whether natural or manmade which reduces or accelerates the reproductive success is evolutionary pressure.

©IJRASET: All Rights are Reserved

For this report, as mentioned in the abstract, it will be the quantitative description of the amount of change occurring in the protein. A 3-D structure of protein consists of amino acid bonds or peptide bonds. (Adhikari and Cheng, 2016) When these peptide bonds are formed, the element of water is removed. Whatever is left is the residue in the amino acids. All amino acids have a 3-letter symbol or a 1 letter symbol. These symbols are essential to differentiate between amino acids. known as mechanism of action, it is defined by a biochemical interaction in which a drug binds to a particular molecular target known as receptors or enzymes. These enzymes have a specific affinity to the drug based on its properties. The action that occurs on the receptor's site is also a result of the affinity between the drug and the enzyme. Another study revealed that proteins have binding pockets which help in determining the structural features of a drug target. These are binding sites where drug and protein sequences interact. These binding pockets have specific properties which affect the protein drug ability. As of now, there are two drugs used for the treatment – oxamniquine and praziquantel. Because oxamniquine is not very efficient in treating all forms of this disease, **praziquantel** is the first option. It is lower in cost and can be used on pregnant women or children alike. Dosage can depend on the person's medical condition, weight, response to treatment, etc. Medication is taken by mouth – with a meal for 3 times a day. In countries where this disease is a major cause of deaths, the WHO runs annual drug shot programs of praziquantel.

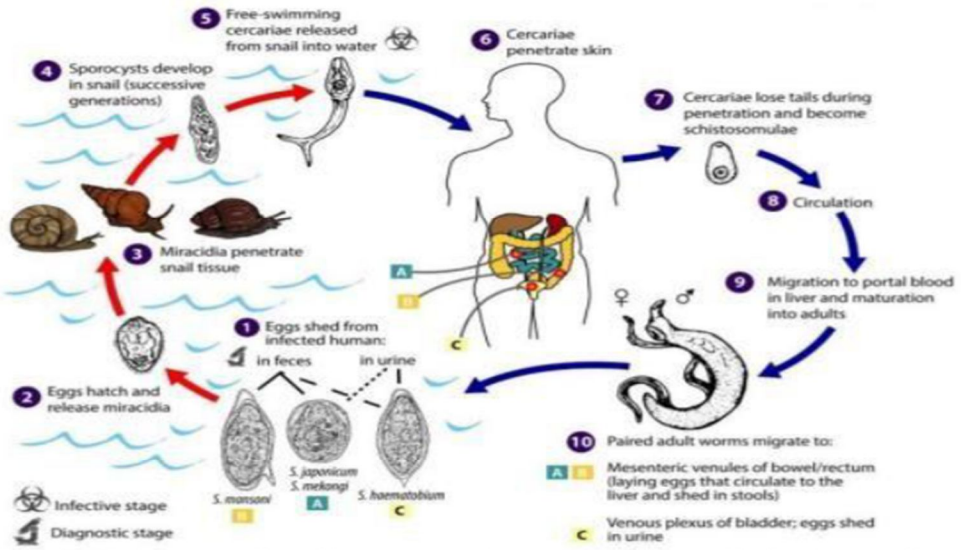


Figure 1: The lifecycle of the Schistosoma requires the involvement of both an intermediate and definitive host in order to be completed. From the starting stage as an egg, they develop to infect freshwater snails to develop and mature further. After exiting the snail into water they can go on to infect their definitive host, which allows them to reach sexual maturity and pair with another worm of the opposite sex to produce eggs which will be shed into the stool and thus continue to the life cycle.

II. MATERIALS

Multiple alignment or multiple sequence alignment are defined as the sequence of 3 or more biological sequences like nucleic acid or proteins.

This report is using a number of methods to analyse the evolutionary pressure on the AA sequences of various species of flatworms.

The first method is Biomart.: As per this article, Biomart is one of the community driven projects which provide a platform to distributed research data and acts as a single point of access. It helps the scientific community and primarily used in biomedical research. It was originated by European Bioinformatics Institute for the Human Genome Project as a data management solution.

Next method is dn/ds: Also known as Ka/Ks ratio. It is used to estimate the balance between neutral mutations – they are neither detrimental nor beneficial for an organism's ability to survive or reproduce, purifying selection and beneficial mutations acting on a homologous protein-coding gene– there are proteins in the same family i.e., having a common ancestor and similar 3-d structure (discussed in phylogenetic trees).

Dn/ds is calculated in the form of ratio between the number of nonsynonymous substitution, i.e., A nucleotide mutation that alters the amino acid sequence of a protein, per nonsynonymous site (Ka) and the number of synonymous substitutions– a mutation which does not alter the amino acid sequence of a protein even after substitution, per synonymous sites (Ks). It is evident that mutation or change occurs after selective process (discussed above).

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

A. Rasmol

It was developed by Roger Sayle in the 90s. It is a computer program used to depict and explore structures of the large biomolecular data or protein mainly found in the protein data bank. It includes a scripting language to perform various functions – changing colours, selecting protein chains, etc.

B. Wormbase

It is a biological database which is available online to study model organism nematodes (roundworms). They are categorized as helminths (parasitic worms). This database is regularly updated and used as an information source by a research community called C. elegans.

C. Dn/ds ratio

Dn/ds gives the ratio of mutations that change the protein structure to the mutations that do not change the structure. This ratio is quite helpful in knowing the type of selection and eventually deciding their closeness to the binding sites. {positive selection is described using the result example in the results section as it needs to be shown using the computed data} (Jeffares et al., 2014)

D. Methods/ Data Replication

The below tools are put in the order of execution

E. Wormbase

Returning to the methodology, the results of protein sequences were downloaded from wormbase (from the archive of known proteins and their databases). Since the aim was searching with TGR, the report used the gene id section for downloading the sequences.

F. Blast

The next step was to run BLAST tool in order to get the database results for a particular sequence given as input. To achieve this, protein sequence was supplied as input and was changed as the sequence option to protein. After running successfully, it gives the result in the form of Excel sheets – which is understandable as it is returning the database information about a sequence.

G. Biomart

Once the results became available, it made sense to - use another tool on wormbase called biomart. The report has discussed biomart (used widely in bioinformatics field by the researchers) above and it helps in returning the FASTA sequences which will be used later on. FASTA sequence is a nucleotide sequence starting with > followed by a unique sequence id.
In biomart, there is a query format to get results. Select GENE, go to gene ID – select transcript stable ID. To fill this, drop down list, go to excel sheet generated by BLAST. In the csv file, select the second column namely subject name and past it in Biomart (Transaction stable ID list). Select Count function to check if the query is as per our requirements. Moving on to the output attributes, select “retrieve sequences.” Then in Sequences select cDNA sequences. Move down to header information, select UniProtKB in gene attributes. The FASTA sequence/ nucleotide sequence files generated using Biomart software. These FASTA sequences need to be aligned.

H. Sequences Used

Since the focus of this report is S mansoni, other flatworm species are chosen based on their functional similarity to S mansoni. It is both an advantage and a disadvantage (discussed in the limitations section). The sequences chosen from uniport or wormbase were S japonicum, S bovis, S haematobium along with S mansoni. They are all blood flukes and hence can show well if mutation near binding sites does impact drug resistance. These sequences are shared in the appendices section (figure 17-20).

I. Clustal Omega

The experiment can either use a tool or align them using a programming language. For example, in python or R – we need to know which data structure to use (mostly string) then each line is parsed (taking into consideration – special characters present in the sequence along with gene ID). Though it is known that not all biologists can code very well, there are some tools which help with the next step.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

These tools are utilized for sequence alignment. Such as MUSCLE, Clustal omega, etc. MUSCLE is a program to create sequence alignment in proteins. It works similarly to Clustal omega. The format of output in both (MUSCLE and Clustal Omega) is PHYLIP. This format is the most popular in bioinformatics tools. The reason for its popularity is that it is easier to read – it has 2 parts. Header describing dimensions of the alignment and next part is the multiple sequence alignment itself. It takes a few seconds to generate the result and even shows if there is any duplicate sequence present. Once the alignments are shown on the screen, we can download the alignment file for further use. The next step is to have computed these alignments using clustal omega in PHYLIP format. These sequences are not similar but of equal length. Equal length gives more accuracy to the results. Also, there are some regions of the alignment which ambiguously aligned i.e. - sometimes, when a computer algorithm is used to align sequences with multiple insertion and deletion, they have gaps in them. Removal of these gaps is subject knowledge. It is good practice to have a replicate sequence and attempt to remove gaps on it. This replicated sequence can show if we can have the desired results after gap-filling. As can be observed from the data above, a number of alignments are grouped together based on a particular homologous criterion – here, it is same species or same classification of flatworm.
They are of same length – http://ugene.net/muscle-multiple alignment-tool/ This webpage shows a method to align sequence without exceptionally good actual programming skills.

J. SNAP Tool

SNAP is used to compare various viruses to each other using the multiple sequence alignments, which it takes as the input. This tool is an essential part of locating evolutionary pressure on multiple sequences. It is a statistical approach by which we can visualize the region of interest in a graphical format as well as detect the values of each AA sequence in terms of syn, non-syn (non-synonymous substitutions), various ratio etc. Since it gives data in a tabular form, they are easy to follow. This table is used to determine regions showing positive selection, which is explained in further sections. (Kryazhimskiy and Plotkin, 2008)

K. PDB

Protein data bank is essential to download the 3-D figure of required protein structure. This is downloaded and opened in Rasmol. This is the structure on which the evolutionary pressure regions would be shown by rasmol.

L. Rasmol

It is a data visualization tool used by bioinformaticians and scientists. It provides visual representation to any protein structure, thus, highlighting the area of research in an effective way. To use rasmol, download the tool. Then, it opens a command prompt along with a black screen. To load image on rasmol, download the sequence from PDB and open in rasmol. Change the display and color depending on the background to make the structure more visible. It uses the table generated in SNAP as input and shows positive selection/ evolutionary pressure regions.

III. RESULTS

```
schistosoma_mansoni       TCAGCGGCGGTAATCGTCTTTAGCAAAACAACTTGTCCATTTTGCAAAAA  27
schistosoma_japonicum     TCAGCGGCCGTAATATTATTTAGCAAGACGACTTGTCCTTATTGCAAAAA  27
schistosoma_haematobium   TCAGCGGCCGTAATATTATTTAGCAAGACGACTTGTCCTTATTGCAAAAA  27

schistosoma_mansoni       GCTAAAGGATGTTTTAGCTGAAGCAAAGATTAAACACGCTACAATTGAAC  28
schistosoma_japonicum     TGTGAAGGATGTTTTGGCTGAAGCAAAGATTAAGCATGCTACAATTGAAC  28
schistosoma_haematobium   TGTGAAGGATGTTTTGGCTGAAGCAAAGATTAAGCATGCTACAATTGAAC  28

schistosoma_mansoni       TGGATCAGTTATCCAATGGTTCGGTTATTCAAAAGGCCTTATCTAGCTTT  29
schistosoma_japonicum     TGGATCAATTATCCAATGGTTCGGCCATTCAGAAGTCCTTAGCCAGCTTC  29
schistosoma_haematobium   TGGATCAATTATCCAATGGTTCGGCCATTCAGAAGTCCTTAGCCAGCTTC  29

schistosoma_mansoni       TCTAAAATTGAAACAGTCCCGCAAATGTTTGTTAGAGGCAAGTTCATTGG  30
schistosoma_japonicum     TCGAAAATTGAAACAGTTCCTCAAATGTTTGTTAGGGGCAAATTCATCGG  30
schistosoma_haematobium   TCGAAAATTGAAACAGTTCCTCAAATGTTTGTTAGGGGCAAATTCATCGG  30

schistosoma_mansoni       CGATTCTAAAGCAGTACTTAATTACCACAATAATAATCAATTGCAGGCGA  31
schistosoma_japonicum     GGATTCTCAGATGGTATTAAAATACCACCGTAATAATGAATTGACGAGTA  31
schistosoma_haematobium   GGATTCTCAGATGGTATTAAAATACCACCGTAATAATGAATTGACGAGTA  31

schistosoma_mansoni       TCGTCAACGAAAATAAGTATGACTATGATCTGATAATCATCGGTGGAGGA  32
schistosoma_japonicum     TTGTCAATGAAAGCAAGTATGACTATGATTTGATAGTTATCGGTGGAGGA  32
schistosoma_haematobium   TTGTCAATGAAAGCAAGTATGACTATGATTTGATAGTTATCGGTGGAGGA  32

schistosoma_mansoni       TCTGGTGGACTCGCTGCTGGAAAGGAGGCAGCCAAATACGGCGCAAAGAC  33
schistosoma_japonicum     TCTGGTGGACTTGCTGCTGGAAAGGAGGCTGCTAAATACGGTGCGAAAAC  33
schistosoma_haematobium   TCTGGTGGACTTGCTGCTGGAAAGGAGGCTGCTAAATACGGTGCAAAAAC  33
```

Figure 9 multiple sequence alignment with sequence numbers (showing only a part of the MSA); these

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Multiple sequence alignment (MSA) as computed using clustal omega tool. As mentioned before, 3 types of flatworms are included for obtaining the MSA. The 3 flatworms chosen were S mansoni, S japonicum, S haemobium. S bovis was also included in the sequences. Their coding sequences were downloaded from Biomart. The sequences which were the input for MSA were pasted in clustal omega textbox along with format type chosen as Phylip interleave – the reason for choosing this format is that is easier to read as there is a header section followed by the multiple sequence alignment itself. The output from clustal omega was an alignment file. These sequences were of equal length to get the appropriate results. There are 600 protein residues amongst the translated part (discussed below). So, they go up to 1800 as there are 3 different flatworm sequences being evaluated. They should be of around same length and not have gaps in between to show that the converted part is error free. This error free MSA is then used in SNAP. In the figure below – Figure 11, green line is synonymous mutations, red is non-synonymous mutations, and the blue line is indel (insertions and deletions). The input for this figure was the MSA generated above. The MSA was pasted into SNAP tool. The output was a graph, codon table, other data about the alignments, etc.

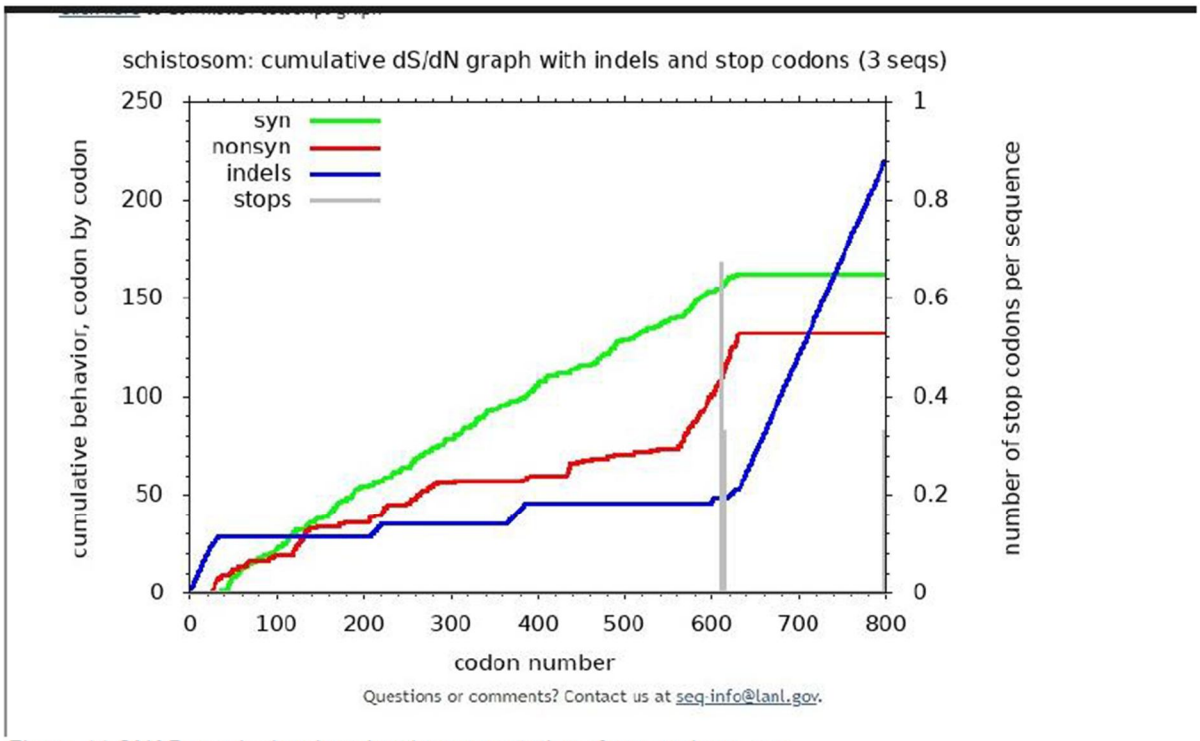
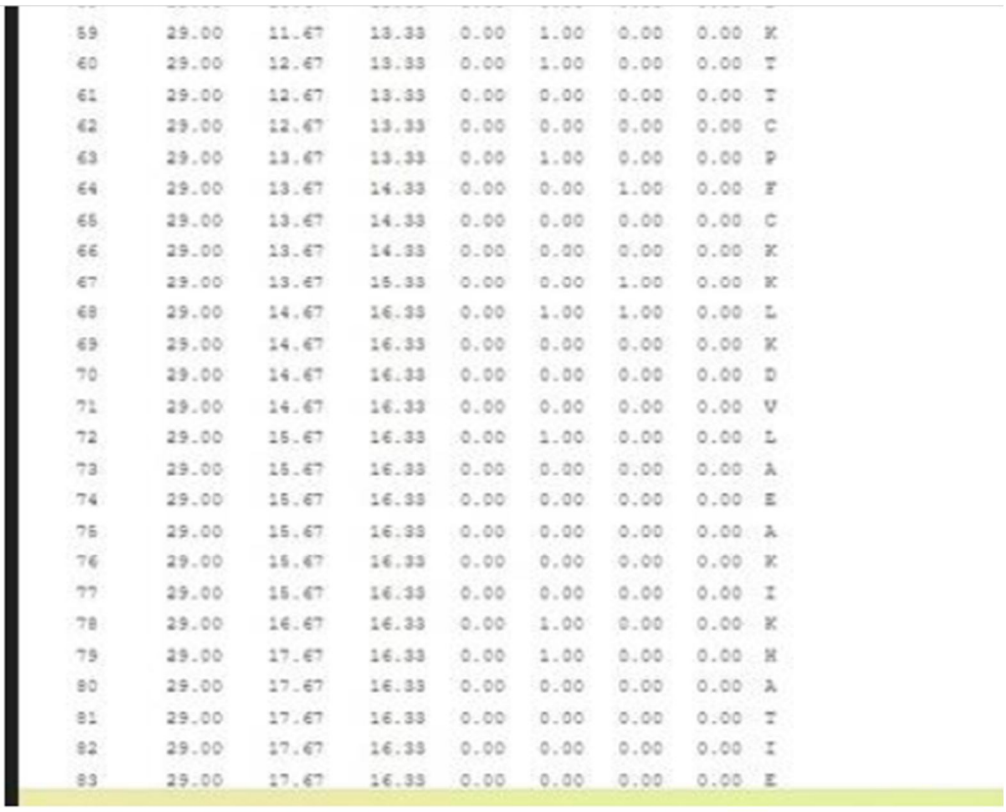


Figure 11 SNAP graph showing visual representation of syn and non-syn

And figure 11 shows, graphically, what the region of interest should be. As evident from the figure, nonsynonymous mutations are rising steeply and the indel becomes flat in this region along with only one codon line – shown by grey line. It indicates that the mutations have altered the AA sequence which is proved by too many indels in the graph. This report chose S japonicum, S mansoni, S haematobium because they are one of the many species causing the disease and are also resistant to praziquantel. Only one codon line shows that there are minimal errors in the output. Hence, this output was fit for analysis. the part which needed analysis is the red and green region because dn/ds ratio is the measure of evolutionary pressure on the protein sequence. And the red and green line signify nonsynonymous and synonymous mutations, respectively. No indels upwards increase in this region signified that there is no rapid insertion or deletion. Rapid insertions and deletions are not good for the results as they corrupt the reading frame. This graph shows an increase in the ratio of nonsynonymous to synonymous mutations – as confirmed by the screenshot of numerical data. This indicates an overall positive selection. the codon data in the form of a table for these multiple sequence alignments. The headings are codon number, syn, non syn, stop codon, aa sequence.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



59	29.00	11.67	13.33	0.00	1.00	0.00	0.00	K
60	29.00	12.67	13.33	0.00	1.00	0.00	0.00	T
61	29.00	12.67	13.33	0.00	0.00	0.00	0.00	T
62	29.00	12.67	13.33	0.00	0.00	0.00	0.00	C
63	29.00	13.67	13.33	0.00	1.00	0.00	0.00	P
64	29.00	13.67	14.33	0.00	0.00	1.00	0.00	F
65	29.00	13.67	14.33	0.00	0.00	0.00	0.00	C
66	29.00	13.67	14.33	0.00	0.00	0.00	0.00	K
67	29.00	13.67	15.33	0.00	0.00	1.00	0.00	K
68	29.00	14.67	16.33	0.00	1.00	1.00	0.00	L
69	29.00	14.67	16.33	0.00	0.00	0.00	0.00	K
70	29.00	14.67	16.33	0.00	0.00	0.00	0.00	D
71	29.00	14.67	16.33	0.00	0.00	0.00	0.00	V
72	29.00	15.67	16.33	0.00	1.00	0.00	0.00	L
73	29.00	15.67	16.33	0.00	0.00	0.00	0.00	A
74	29.00	15.67	16.33	0.00	0.00	0.00	0.00	E
75	29.00	15.67	16.33	0.00	0.00	0.00	0.00	A
76	29.00	15.67	16.33	0.00	0.00	0.00	0.00	K
77	29.00	15.67	16.33	0.00	0.00	0.00	0.00	I
78	29.00	16.67	16.33	0.00	1.00	0.00	0.00	K
79	29.00	17.67	16.33	0.00	1.00	0.00	0.00	H
80	29.00	17.67	16.33	0.00	0.00	0.00	0.00	A
81	29.00	17.67	16.33	0.00	0.00	0.00	0.00	T
82	29.00	17.67	16.33	0.00	0.00	0.00	0.00	I
83	29.00	17.67	16.33	0.00	0.00	0.00	0.00	E

For the unversed, codon is a sequence of 3 consecutive nucleotide in a DNA or RNA molecule that codes for a specific amino acid (aa). Here, syn means that the meaning of sequence of DNA is unchanged when it is translated in AA. Each DNA triplet is translated into amino acid for the protein sequence. the codon data from SNAP which is computed in the form of a table (shown in table 1 and 2) was exported into excel worksheet.

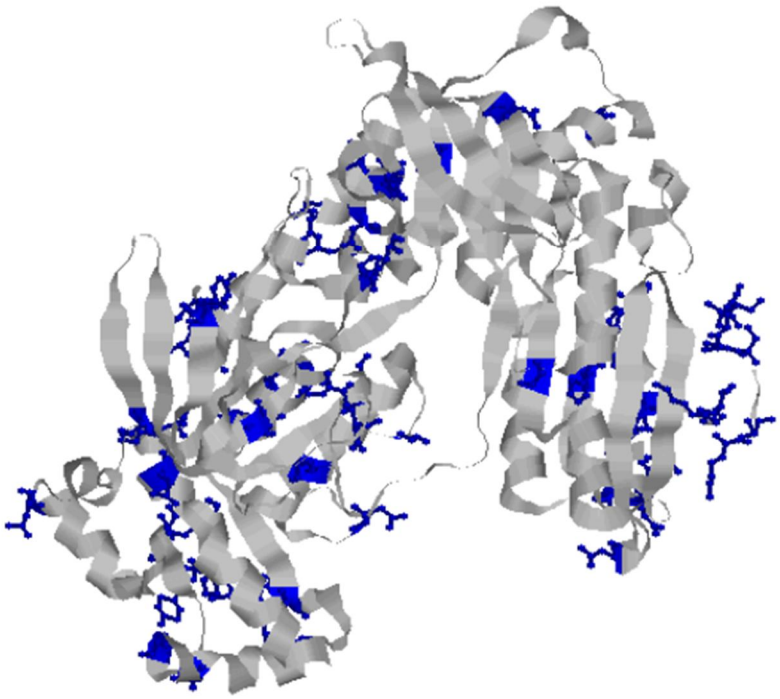
	codon	indel	syn	nonsyn	indel	syn	nonsyn	stopc	aa	dn/ds ratio
2	25	24.50	0.00	0.50	0.50	0.00	0.50	0.00	-	>1
3	27	25.50	0.00	1.50	0.50	0.00	1.00	0.00	-	>1
4	28	26.00	0.00	2.50	0.50	0.00	1.00	0.00	-	>1
5	29	26.50	0.08	3.92	0.50	0.08	1.42	0.00	-	>1
6	30	27.00	0.17	5.33	0.50	0.08	1.42	0.00	-	>1
7	31	27.50	0.17	6.33	0.50	0.00	1.00	0.00	-	>1
8	32	28.00	0.17	6.83	0.50	0.00	0.50	0.00	-	>1
9	33	28.50	0.17	7.33	0.50	0.00	0.50	0.00	-	>1
0	34	29.00	0.17	7.83	0.50	0.00	0.50	0.00	-	>1
1	38	29.00	1.17	8.83	0.00	0.00	1.00	0.00	I	>1
2	50	29.00	8.17	11.83	0.00	0.00	1.00	0.00	E	>1
3	56	29.00	10.67	13.33	0.00	0.50	1.50	0.00	V	>1
4	64	29.00	13.67	14.33	0.00	0.00	1.00	0.00	F	>1
5	67	29.00	13.67	15.33	0.00	0.00	1.00	0.00	K	>1
6	94	29.00	20.67	17.33	0.00	1.00	0.00	0.00	Q	>1
7	121	29.00	31.33	22.67	0.00	0.67	2.33	0.00	A	>1
8	124	29.00	32.33	24.67	0.00	0.00	1.00	0.00	N	>1
9	127	29.00	32.33	26.67	0.00	0.00	2.00	0.00	N	>1
0	130	29.00	32.33	27.67	0.00	0.00	1.00	0.00	Q	>1
1	132	29.00	32.33	29.67	0.00	0.00	2.00	0.00	Q	>1
2	133	29.00	33.00	32.00	0.00	0.67	2.33	0.00	A	>1

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

After pasting it in excel, another column was added namely dn/ds ratio. This column calculated the ratio based on individual syn and non-syn values. As mentioned in table 1, there are cumulative and individual syn and non-syn values given. This is computed for each codon number. They are mainly divided/ categorized as 1, >1, <1 or 0. Interpretation of the ratio value is given below. For now, the ratio >1 is highlighted in yellow to show regions of evolutionary pressure in the protein structure. It gives a clear understanding of the location through the codon number.

IV. DISCUSSION INCLUDING LIMITATIONS

The result section details about the findings as well as the procedure followed to reach those findings. The MSA shows the alignment of these flatworms, which is later used in SNAP tool to generate a graph. With the help of this graph, it is inferred that no errors are present in the region of interest. The region of interest is where the syn and non syn values show changes. Then the codon table is used to determine the individual selection. After analyzing them and most importantly the graph, it has been inferred that evolutionary pressure might have a role to play in drug resistance. The regions showing translation and changes in syn, non syn values are of utmost importance. Since stop codons are not an issue in the results, all the translated portions are worthy of analysis. As shown in various images and tables in the results section, the regions of interest to generate a conclusion are with ratio >1 and they are well highlighted. Protein data bank is a database for large molecular structures like proteins. It shows 3-D representation of these molecules. Below is a PDB figure of 3D structure of the selected protein. Going through the technical part of the report, thioredoxin glutathione reductase is an enzyme/protein associated with resistance against praziquantel. In PDB, search the 2x8h in the entry. After opening the specific page, there is a dropdown menu on the right side of the page. Select FASTA sequence, download it and open it in rasmol. Figure 14 shows the 3d structure of the protein as shown in PDB. Once this PDB file is opened in Rasmol, Another way to detect sites of evolutionary pressure is through RASMOL. The basis of visualizing in rasmol for this project is the data in Table 3 and 4 along with tables 5,6,7 and the highlighted region. These regions clearly show the ratio >1 and, hence, are the locations where mutation in the protein structure takes place. To visualize, the commands are typed on the Rasmol command prompt.



As shown in Figure 15, the entire structure is selected. It is generated by typing select backbone on the rasmol command prompt. It selects the entire structure. After that, to select evolutionary pressure sites based on the tables 3-7, those specific codon number values are selected. As is evident, most of the locations of evolutionary pressure (figure 15) are located near the binding sites (as discussed in introduction section).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

When binding sites cluster together they form a binding pocket which determines the properties of a protein structure. Figure 16 shows a few regions where a few pocket of binding sites have formed, this helps in showing that binding pocket properties are indeed influencing drug resistance because they change the functionality of the protein.

Since this report only analyses the S mansoni flatworm, it is not certain if the same experiment will yield similar results for other types of flatworms or even other types of parasites. Due to time constraint, for multiple sequence alignment, the choice of flatworms was narrowed down to 3-4 which resemble the structure of 2x8h protein, it is not clear if the results would be satisfactory if other types of flatworms are used (not having structure like 2x8h). As per the cited paper, sometimes the dn/ds ratio are not a true indicator of positive or negative selection. This means further investigation is needed in some cases. (Rahman et al., 2021)

V. CONCLUSION AND FUTURE WORKS

The aim of the project was to find regions of evolutionary pressure in S mansoni protein sequence which helps it in mutating and these mutations help it to evade a drug called praziquantel. For this, sequences are downloaded from biomart, then blasted. These sequences are then aligned using clustal omega. Also, DNA sequence is converted into AA sequence, and it is visualized in figure 10. DNA sequences of a few flatworms were chosen based on their similarity with S mansoni to be used for MSA. This MSA is compared using SNAP tool. To reiterate, the graph generated by snap tool along with the various tables show that there are very few stop codons which means no or less signs of error in the data. The translated part of the sequence is shown using a letter in AA field, accompanied by syn and non syn values. Where syn and non syn values are greater than 0 – it shows mutations. But values between 0 and 1 do not show many mutations. It is the values greater than 1 of non syn in particular which are of interest to this experiment, as demonstrated in the tables 3 and 4. A non syn value greater than 1 shows high dn/ds ratio, as the yellow highlighted region in tables 3,4,5,6,7. The locations in a protein structure showing high dn/ds ratio determine the regions of evolutionary pressure. These regions are the ones which change the mutation of the protein structure significantly for the flatworm to resist the drug called praziquantel. They are visually represented using rasmol tool. With the help of the above-mentioned tables, it was easier to determine the translated regions which show positive selection. Also, their codon number and aa symbol helped in locating the said sites into rasmol. And as it was expected, it also shows that the proximity of the evolutionary pressure regions to the binding sites influences their resistance towards the drug Praziquantel. The reason for it is that if the mutation occurs near a binding site, the drug which binds to it on that specific space will react differently to the structure. Hence, the drug will lose its effectiveness in countering the mutated structure. Lastly, as mentioned in the introduction section, the gaps in knowledge were related to the lack of knowledge of which AA sequence are under evolutionary pressure. This has been successfully addressed in this report as well by determining the aa residues of positive selection codon numbers in tables 3 to 7. Biophysical structure change leads to mutation which help a flatworm in drug resistance, as demonstrated by this report. There is no limit to further research, some of the future projects are mentioned below. For example - to understand the biophysical effects on natural selection and, eventually evolutionary pressure, protein mutation effects are being surveyed. Their effects are visualized on spatial representation of polypeptide chains. Methods are being developed to exploit sequence based evolutionary information. This will greatly help in predicting biophysical behaviors in protein. The biophysical behavior (which changes the protein structure) plays a big role in determining drug resistance

A. Appendices

REFERENCES

[1] Angelucci, F., Dimastrogiovanni, D., Boumis, G., Brunori, M., Miele, A.E., Saccoccia, F. and Bellelli, A. (2010). Mapping the Catalytic Cycle of Schistosoma mansoni Thioredoxin Glutathione Reductase by Xray Crystallography*. Journal of Biological Chemistry, [online] 285(42), pp.32557–32567. doi:https://doi.org/10.1074/jbc.M110.141960.

[2] Eweas, A.F. and Allam, G. (2018). Targeting thioredoxin glutathione reductase as a potential antischistosomal drug target. Molecular and Biochemical Parasitology, 225, pp.94–102.doi:https://doi.org/10.1016/j.molbiopara.2018.09.004.

[3] Fata, F., Silvestri, I., Ardini, M., Ippoliti, R., Di Leandro, L., Demitri, N., Polentarutti, M., Di Matteo, A., Lyu, H., Thatcher, G.R.J., Petukhov, P.A., Williams, D.L. and Angelucci, F. (2021). Probing the Surface of a Parasite Drug Target Thioredoxin Glutathione Reductase Using Small Molecule Fragments. ACS infectious diseases, [online] 7(7), pp.1932–1944. doi:https://doi.org/10.1021/acsinfecdis.0c00909. Kryazhimskiy, S. and Plotkin, J.B. (2008). The Population Genetics of dN/dS. PLoS Genetics, [online]4(12), p.e1000304. doi:https://doi.org/10.1371/journal.pgen.1000304.

[4] Littlewood, D.T.J. (2006). The evolution of parasitism in flatworms. Parasitic flatworms: molecular biology, biochemistry, immunology and physiology, pp.1–36. doi:https://doi.org/10.1079/9780851990279.0001. Lyu, H., Petukhov, P.A., Banta, P.R., Jadhav, A., Lea, W.A., Cheng, Q., Arnér, E.S.J., Simeonov, A., Thatcher, G.R.J., Angelucci, F. and Williams, D.L. (2020). Characterization of Lead Compounds Targeting the Selenoprotein Thioredoxin Glutathione Reductase for Treatment of Schistosomiasis. ACS infectious diseases, [online] 6(3), pp.393–405. doi:https://doi.org/10.1021/acsinfecdis.9b00354.

[5] Mustacich, D. and Powis, G. (2000). Thioredoxin reductase. The Biochemical Journal, [online] 346 Pt 1(Pt1), pp.1–8. Available at: https://pubmed.ncbi.nlm.nih.gov/10657232/.

[6] Rahman, S., Kosakovsky Pond, S.L., Webb, A. and Hey, J. (2021). Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. Proceedings of the National Academy of Sciences, 118(20). doi:https://doi.org/10.1073/pnas.2023575118.

[7] Schmidt, C.L.A. and Thomas, C.C. (1939). The Chemistry of the Amino Acids and Proteins. Soil Science, [online] 47(2), p.168. Available at: https://journals.lww.com/soilsci/citation/1939/02000/the_chemistry_of_the_amino_acids_and_proteins.12.aspx [Accessed 8 Mar. 2023].

[8] Sievers, F. and Higgins, D.G. (2017). Clustal Omega for making accurate alignments of many protein sequences. Protein Science, 27(1), pp.135–145. doi:https://doi.org/10.1002/pro.3290.

[9] Sikosek, T. and Chan, H.S. (2014). Biophysics of protein evolution and evolutionary protein biophysics. Journal of the Royal Society, Interface, [online] 11(100), p.20140419. doi:https://doi.org/10.1098/rsif.2014.0419.

[10] Thompson, J.D., Gibson, Toby.J. and Higgins, D.G. (2002). Multiple Sequence Alignment Using ClustalW and ClustalX. Current Protocols in Bioinformatics, 00(1), pp.2.3.1–2.3.22. doi:https://doi.org/10.1002/0471250953.bi0203s00.

[11] Jeffares, D.C., Tomiczek, B., Sojo, V. and dos Reis, M. (2014). A Beginners Guide to Estimating the Nonsynonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome. Methods in Molecular Biology, pp.65–90.

[12] Chu, D. and Wei, L. (2019). Nonsynonymous, synonymous and nonsense mutations in human cancer related genes undergo stronger purifying selections than expectation. BMC Cancer, 19(1).

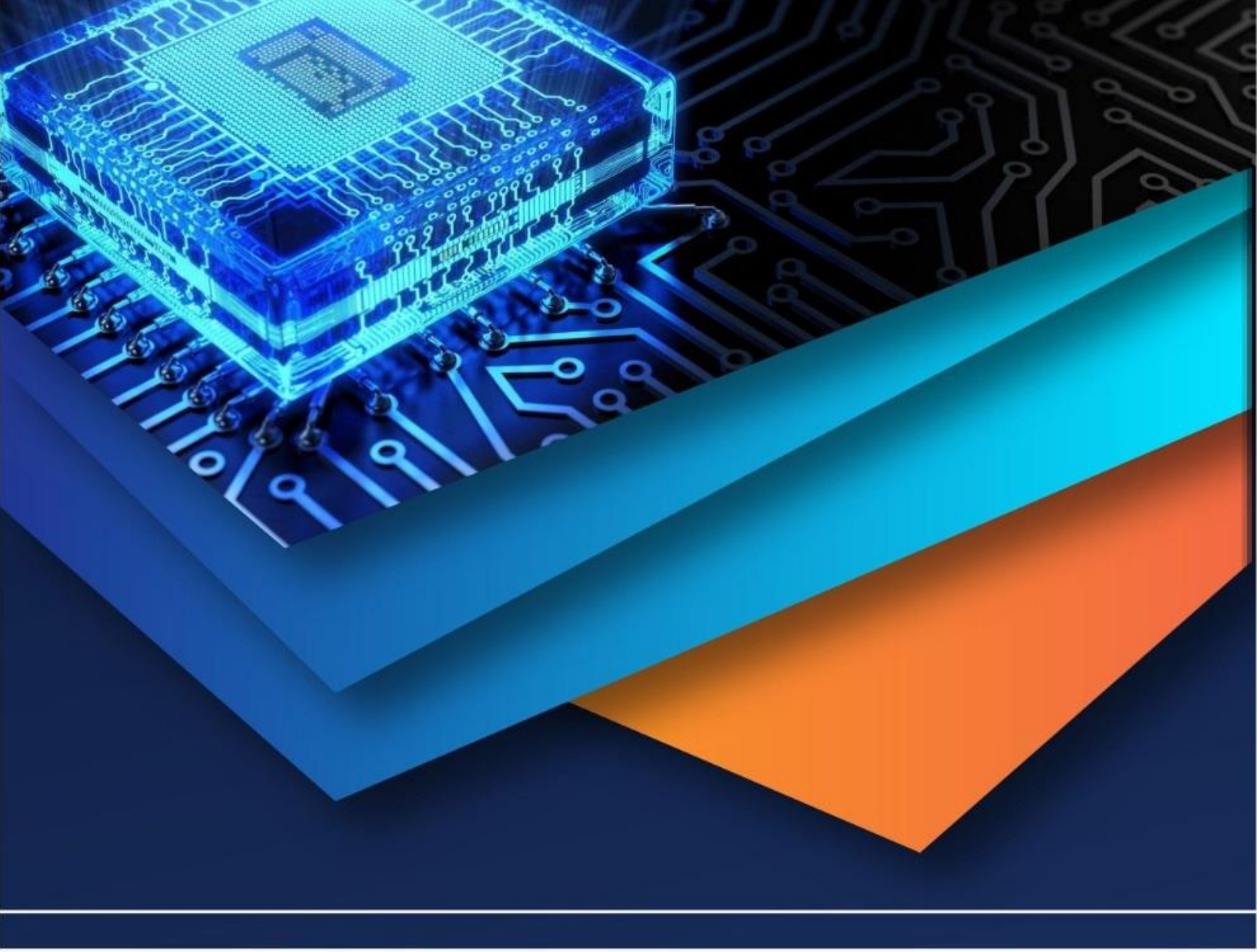
[13] Wagner, A. (2007). Rapid Detection of Positive Selection in Genes and Genomes Through Variation Clusters. Genetics, [online] 176(4), pp.2451–2463. doi:https://doi.org/10.1534/genetics.107.074732.

[14] Adhikari, B. and Cheng, J. (2016). Protein Residue Contacts and Prediction Methods. Methods in Molecular Biology, pp.463–476. doi:https://doi.org/10.1007/978-1-4939-3572-7_24.

[15] Mattos, C. and Ringe, D. (1996). Locating and characterizing binding sites on proteins. Nature Biotechnology, 14(5), pp.595–599. doi:https://doi.org/10.1038/nbt0596-595.

[16] www.ncbi.nlm.nih.gov. (n.d.). 2X8H: Thioredoxin glutathione reductase from Schistosoma mansoni in complex with GSH. [online] Available at: https://www.ncbi.nlm.nih.gov/Structure/pdb/2X8H.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)