International Journal of Entomology Research
ISSN: 2455-4758; Impact Factor: RJIF 5.24
Received: 03-10-2020; Accepted: 18-10-2020; Published: 08-11-2020
www.entomologyjournals.com
Volume 5; Issue 6; 2020; Page No. 57-65

Structural functional insights into Earthworm Lysenin as potential antimicrobial agent

Shyamasree Ghosh
School of Biological Sciences, National Institute of Science Education and Research (NISER), Bhubaneswar, Odisha, India

Abstract
The innate immune system in primitive organism of earthworms belonging to the invertebrate Phylum of Annelida, exhibits advanced functional features and plays role in conferring immunity. The innate immune system of earthworms, consist of phenoloxidase system (POS), pattern recognition rectors (PRRs), Toll like receptors (TLRs), antimicrobial peptides (AMPs) coelom cytolytic factor and lysenin that confer immunity to the soil dwelling earthworms and is continuously exposed to pathogens and pollutants, however detailed mechanism of their function remains largely unknown. In this article we have (i) studied the wild type lysenin from *Eisenia fetida* (*E fetida*) (ii) identified the binding pocket and (iii) studied the binding of wild-type lysenin from *E fetida* with membrane protein of Shigella *flexneri (S flexneri)* by *insilico* methods of protein-protein docking. We have tried to understand the implications of lysenin in targeting pathogen by using the Cluspro 2.0 and Schrodinger protein-protein docking tool. This is the first report of structure and functional *insilico* based study of lysenin binding to bacterial surface protein, providing an insight of the mode of action of lysenin in conferring immunity against *S flexneri* infection with a potential application as an antibacterial agent.

Keywords: Lysenin, innate immune system, pattern recognition receptors (PRRs), *Shigella flexneri*

Introduction
Earthworms are exposed to the harsh conditions of the soil in the polluted environment and are continually challenged by microbial pathogens of the environment. However the earthworm immune system is robust enough to combat the invading pathogens. The coleomic fluid of earthworm houses vital components of the earthworm innate immune system comprising of the phenoloxidase system (POS), pattern recognition rectors (PRRs), Toll like receptors (TLRs), antimicrobial peptides (AMPs, Ghosh 2019) [7], coelom cytolytic factor (CCF, Ghosh 2020) [8] and lysenin; Bilej *et al*, (2000-2013) [2], Ghosh (2018) [9], Prochazkova *et al*, (2017) [25].
Lysenin is known for its antimicrobial role and is known to confer immunity; Swiderska (2017) [28], Kobayashi *et al*, (2004) [17] to the earthworms against pathogens.
In this study we have performed *insilico* analysis characterising the wild type lysenin from *Eisenia fetida (E fetida*), predicting its structural features. To study its antimicrobial role, we performed the *insilico* docking studies with the membrane protein from *Shigella flexneri (S flexneri*) providing an insight into its important biological function with a probable pharmocological application in combating *Shigella* infection which is a gram-negative bacteria, belonging to serogroup B, inhabiting the epithelium of the colon causing severe inflammation and tissue destruction in humans, crossing the epithelial barrier causing diarrhoea and shigellosis leads to more than 1.1 million deaths and over 165 million cases reported annually, in humans; Jennison *et al*, (2004) [16]. Although the disorder is treated by antibiotics, the rapidly rising antibiotic resistance is a major problem in targeting this infection.
We have tried to understand the protein-protein interaction by using the Cluspro 2.0 and the Schrodinger protein-protein docking tools. This study finds importance in Establishing the role of lysenin in targeting *Shigella* as a potent antimicrobial molecule, in the light of search for new antimicrobials in targeting bacterial infection.

2.1 Materials and Methods
Wild-type lysenin with PDB Code 3ZXD and *Shigella* membrane protein with PDB Code 4Q35 with sequence as in Table 1 were selected for the study.

2.2. Study of protein properties
The characteristics of the protein was studied using the Protparam; Gasteiger (2005) [6], from the Expasy tools.

2.3 Detection of phosphorylation sites
The NetPhos 3.1 server was used to predict the serine, threonine or tyrosine phosphorylation sites in eukaryotic proteins using neural networks. Both generic and kinase specific predictions were performed. Predictions are made for the 17 kinases including ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38MAPK; Blom *et al*, (2004) [3].

2.4 Prediction of tyrosine sulfanition
Tyrosine sulfation is an important post-translational modification (PTM) of proteins undergoing the secretory pathway. We used the Expasy Sulfinator tool; Monigatti *et al*, (2002) [22] to predict the probable sulfination of tyrosine residues in the wild type lysenin

2.5. Prediction of *O*-GlcNAc sites
We used the Expasy tool and the YinOYang WWW server for prediction of O-ß-GlcNAc attachment sites; Gupta and Brunak, (2002) [3] on the wild type lysenin, protein. This server use NetPhos for predicting possible phosphorylated sites and thereby identifying "Yin-Yang" sites.

2.6 Finding the computed surface topology of the protein
We used the online tool CASTp, Computed Atlas of Surface Topography of proteins, tools to detect the binding pocket or empty concavities on a protein surface here wild type lysenin into which solvent (probe sphere 1.4 A) can gain access, i.e., these concavities have mouth openings connecting their interior with the outside bulk solution. Shallow depressions are excluded from the calculation by the methodology. A pocket is distinguished from a shallow depression by the fact of consideration that among the infinite number of possible cross sections of a pocket, at least one is larger than the mouth opening of the pocket; Tian *et al*, (2018) [29].

2.7. Protein protein docking tools
Sites of expasy (https://www.expasy.org/proteomics) and Cluspro; Kozakov *et al*, (2017, 2013) [19] was used to study protein-protein docking of wild-type lysenin with PDB Code 3ZXD with the bacterial membrane protein from *S flexneri* 4Q35, Qiao *et al*, (2014) for peptide docking.
2.8. Maestro interphase of Schrödinger was used to further confirm the binding. The protein structure 3ZXD (wild-type lysenin) and the 4Q35 (bacterial membrane protein from *S flexneri*) was processed using the Maestro interphase of Schrödinger (https://www.schrodinger.com/maestro) using protein preparation wizard (https://www.schrodinger.com/protein-preparation-wizard).
The missing hydrogens were added, the disulfide bonds were created, missing side chains and residues were added and water molecules were deleted, the ionization state for the amino acid was predicted finally the protein structure was minimized. The protein-protein docking was carried out using the PIPER module in Schrödinger (https://www.schrodinger.com/science-articles/biologics-design). The rigid body docking was performed using the default parameters, where finally 30 poses of the protein-protein complex was saved. The pose with highest cluster was considered as best pose. Further the PIPER pose energy and PIPER pose score, are calculated represents the interaction energy between two proteins with the following expression:
E = w1Erep + w2Eatt + w3Eelec + w4EDARS,
Where wn are weights, Erep and Eatt are the repulsive and attractive contributions to the van der Waals interaction energy, and Eelec is an electrostatic energy term. The sum of the first three (weighted) terms is reported as PIPER pose energy. EDARS is a pairwise structure-based potential constructed by the 'decoys as the reference state' (DARS) approach, and it primarily represents desolvation contributions. The fourth term (weighted) is reported as the PIPER pose score (Kozakov *et al.,* 2017) [19]. The complex was analyzed using the protein interaction analyzer tool. The tool lists out all the non-bonded interactions and the Van der Waals shape complementarily (Lawrence *et al.*, 1993) and the fraction of the solvent-accessible surface area of the residue that is buried by the interaction with residues in other protein. The binding energy was calculated further by running the Prime-MMGBSA.The binding energy was calculated according to the equation:
DG bind = E_complex (minimized) - E_ligand (minimized) -E_receptor (minimized). DG = -112.4

Results
The earthworm (Fig. 1A) lysenin wild type (Table 1) and the binding pocket was predicted as in Fig. 1B.Lysenin were studied for physicochemical characteristics in indicative of a stable protein (Table 2) and 43 predicted sites for phosphorylation (Table 2). The tyrosine sulfanition residue is predicted at position 281 in the site PATNVYCL-DKREDKWI (Table 3).Amino acids 20, 187, 209, 245, 291 in chains A, B, C, D, are crucial for binding sphingomyelin and inducing hemolysis activity amino acid (aa) at 248 is predicted as glycosylation site, while disulphide bond is predicted at 272-283 (Table 4).
Recently reports have highlighted the important role of glycan-glycan interactions in mediating binding of *Shigella* to host epithelial cells; Belotserkovsky *et al*, (2018) [1]. We, in this study report the predicted *O-ß*-GlcNAc attachment sites in the wild type lysenin including 32S, 54T, 77S, 110S, 143S, 149S, 149T, 150S, 182S, 190T, 218S, 225S, 237T with most predictability in sites 54T and 110S (Table 3) which could play dominant role in the lysenin wild type-Shigella membrane protein interaction.
From the surface topology analysis by CASTp of lysenin, monomer, it revealed while mutations in site 20 is predicted to confer property of loss of ability to bind sphingomyelin, and to induce hemolysis, mutagenesis at aa 187 is indicative of loss of ability to bind sphingomyelin, and induces hemolysis only at high concentration. Mutagenesis at aa 116, the protein is not expressed, at 206, mutagenesis site predicted does not affect binding, and has little loss of hemolytic activity, mutagenesis of aa at 245 and 291 confers loss of ability to bind sphingomyelin, and to induce hemolysis (Table 4).
We report the binding of lysenin wild type with *Shigella* membrane protein by docking studies using the Cluspro peptide docking tools (Fig. 2), thus offering promises as an activity as an antibacterial agent in *Shigella* infection.
ClusPro is a web based server is a direct docking method and analyses and creates model for two interacting proteins. Docking methods can be classified as direct or template-based. Based on thermodynamics, direct methods attempt to find the structure of the target complex located at the minimum of Gibbs free energy in the conformational space, and thus require a computationally feasible free energy evaluation model and an effective minimization algorithm.
ClusPro performs three computational steps to predict the model of interaction using regression algorithm. These steps include (a). The rigid body docking by sampling billions of conformations, (b) root-mean-square deviation (RMSD) based clustering of the 1000 lowest energy structures generated to find the largest clusters that will represent the most likely models of the complex, and (c) refinement of selected structures using energy minimization. Rigid body docking step uses PIPER algorithm based on the Fast Fourier Transform (FFT) based correlation approach.
ClusPro server is based on PIPER which performs the sampling. The center of mass of the receptor is fixed at the origin of the coordinate system, and the possible rotational and translational positions of the ligand are evaluated at the given level of discretization. The rotational space is sampled on a sphere-based grid that defines a subdivision of a spherical surface in which each pixel covers the same surface area as every other pixel.

PIPER the following equation to represent the interaction energy between the two interacting proteins:

$E = w_1 E_{rep} + w_2 E_{att} + w_3 E_{elect} + w_4 E_{DARS}$

Where E_{rep} is the repulsive contribution and E_{att} is the attractive contribution to the Van der Waals interaction energy, E_{elect} is the electrostatic energy and E_{DARS} is the pair wise structure based contribution i.e. the free energy change due to the removal of the water molecule from the interface. w1, w2, w3, w4 are the weight factors optimally determined for the docking problem in hand. The server generates four sets of models using the scoring schemes called (1) balanced, (2) electrostatic-favored, (3) hydrophobic-favored, and (4) van der Waals + electrostatics.

The results of protein binding between Lysenin and *Shigella flexneri* shows an estimated correlation:

$E = 0.40 E_{rep} + -0.40 E_{att} + 600 E_{elect} + 1.00 E_{DARS}$

With w_1=+0.40, w_2=-0.4 signifying softening of the van der Waals forces, w_3 = +600 is the balanced option as charges are measured in multiple electron charges and the distance is measured in A^0, and since E_{DARS} is scaled to the magnitudes of protein-protein binding free energies, w_4=+1.0 is the "neutral" choice.Using the Maestro Schrodinger, the pose with highest cluster size which was ranked top exhibited a pose score of -504.8 kcal/mol and pose energy of -1139.9 kcal/mol (Figure 2).

The interacting residues with a cut off value of surface complementarily selected at 0.65 are detailed in Table 5. The Glu391 of Shigella membrane protein with PDB Code 4Q35 protein formed salt bridge with the Lys21 of Wild-type lysenin with PDB Code 3ZXD ; there were hydrogen bond observed between 291Tyr of 4Q35 with the Arg29 of 3ZXD, Gln218 of 4Q35 and Glu85 of 3ZXD, Gln291 of PDB Code 4Q35 with Glu85 of 3ZXD. Other surface complementing hydrophobic interactions are listed in Table 5. The dG of binding calculated for the complex was found to be -112.4 kcal/mol.

Table 1: Sequences

Name and PDB Code of Protein	Sequence
Wild-type lysenin>3ZXD:A|PDBID|CHAIN|SEQUENCE	SAKAAEGYEQIEVDVVAVWKEGYVYENRGSTSVDQKITITKGMKNVNSETRTVTATHSIGSTISTGDAFEIGSVEVSYSHSHEESQVSMTETEVYESKVIEHTITIPPTSKFTRWQLNADVGGADIEYMYLIDEVTPIGGTQSIPQVITSRAKIIVGRQIILGKTEIRIKHAERKEYMTVVSRKSWPAATLGHSKLFKFVLYEDWGGFRIKTLNTMYSGYEYAYSSDQGGIYFDQGTDNPKQRWAINKSLPLRHGDVVTFMNKYFTRSGLCYDDGPATNVYCLDKREDKWILEVVGLVPRGSGHHHHHH
>4Q35:A|PDBID|CHAIN|SEQUENCE	MKKRIPTLLATMIATALYSQQGLAADLASQCMLGVPSYDRPLVQGDTNDLPVTINADHAKGDYPDDAVFTGSVDIMQGNSRLQADEVQLHQKEAPGQPEPVRTVDALGNVHYDDNQVILKGPKGWANLNTKDTNVWEGDYQMVGRQGRGKADLMKQRGENRYTILDNGSFTSCLPGSDTWSVVGSEIIHDREEQVAEIWNARFKVGPVPIFYSPYLQLPVGDKRRSGFLIPNAKYTTTNYFEFYLPYYWNIAPNMDATITPHYMHRRGNIMWENEFRYLSQAGAGLMELDYLPSDKVYEDEHPNDDSSRRWLFYWNHSGVMDQVWRFNVDYTKVSDPSYFNDFDNKYGSSTDGYATQKFSVGYAVQNFNATVSTKQFQVFSEQNTSSYSAEPQLDVNYYQNDVGPFDTRIYGGQAVHFVNTRDDMPEATRVHLEPTINLPLSNNWGSINTEAKFLATHYQQTNLDWYNSRNTTKLDESVNRVMPQFKVDGKMVFERDMEMLAPGYTQTLEPRAQYLYVPYRDQSDIYNYDSSLLQSDYSGLFRDRTYGGLDRIASANQVTTGVTSRIYDDAAVERFNISVGQIYYFTESRTGDDNITWENDDKTGSLVWAGDTYWRISERWGLRGGIQYDTRLDNVATSNSSIEYRRDEDRLVQLNYHYASPEYIQATLPKYYSTAEQYKNGISQVGAVASRPIADRWSIVGAYYYDTNANKQADSMLGVQYSSCCYAIRVGYERKLNGWDNDKQHAVYDNAIGFNIELRGLSSNYGLGTQEMLRSNILPYQNTLMRYLATLLLSLAVLITAG

Table 2: Physico -Chemical parameters of AMPs[@] and Predicted Phosphorylation site[$$] of wild type lysenin

No. of amino acids	M. wt	pI	-R (Asp+ Glu)	+R (Arg+ Lys)	Ext. coefficient (EC)[#]	Instability index (II)[##]	Aliphatic index (AI)[###]	GRAVY[####]	Predicted phosphorylation site
309	34799.10	6.35	38	34	58455	35.10 stable	75.95	-0.459	SAKAAEGYEQIEVDVVAVWKEGYVYENRGSTSVD QKITITKGMKNVNSET # 50 RTVTATHSIGSTISTGDAFEIGSVEVSYSHSHEESQV SMTETEVYESKVI # 100 EHTITIPPTSKFTRWQLNADVGGADIEYMYLIDEVT PIGGTQSIPQVITS # 150 RAKIIVGRQIILGKTEIRIKHAERKEYMTVVSRKSWP AATLGHSKLFKFV # 200 LYEDWGGFRIKTLNTMYSGYEYAYSSDQGGIYFDQ GTDNPKQRWAINKSL # 250 PLRHGDVVTFMNKYFTRSGLCYDDGPATNVYCLD KREDKWILEVVGLVPR # 300 GSGHHHHHH # 350 %1 S.....Y.....Y.Y....STS.....T.....S. # 50 %1 .T.....S...S...ST.....SYS.S...S...S.T.T...Y.S... # 100 %1...T.....T.....T.....T. # 150 %1 .....T.....Y.T...S...S....T..... # 200 %1 .Y.....T.....S...Y.Y.S.....S. # 250 %1 .....Y..... # 300 %1 .S.....

[#] Extinction coefficients expressed in M^{-1} cm^{-1}, at 280 nm measured in water reveals the concentration of Cys, Trp and Tyr amino acid. The

computed Extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution.
[##] The instability index (II) is indicative of stability of protein in a test tube. A value smaller than 40 is predicted as stable, above 40 indicates the protein may be unstable.
[###]Aliphatic index is indicative of stability for a wide temperature range.
[####]The Grand Average hydropathy (GRAVY) value is measured as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. This low range of value indicates the possibility of better interaction with water.

Table 3: Prediction of tyrosine sulphination site

Predicted tyrosine sulphination site[@@]				Residue *O*-GlcNAc result	
Protein / sequence name	Position	E-value	Sequence	32 S +	
3ZXD	281	[46]	PATNVYCL-DKREDKWI	54 T +++	
			+ + Y+ + +++ +	77 S +	
				110 S ++	
				143 S +	
				149 T +	
				150 S +	
				182 S +	
				190 T +	
				218 S +	
				225 S +	
				237 T +	

[@@]E cut off value is 55 and sulphated tyrosines are detected 1 of 17

Table 4: Surface Topography Analysis of wild type lysenin (3ZXD)

Feature	Position(s)	Description
Site	A: 20	Crucial for binding sphingomyelin and inducing hemolysis
Site	A: 187	Crucial for binding sphingomyelin and important for inducing hemolysis
Site	A: 209	Important for activity
Site	A: 245	Crucial for binding sphingomyelin and inducing hemolysis
Site	A: 291	Crucial for binding sphingomyelin and inducing hemolysis
glycosylation site	A: 248	N-linked (GlcNAc...) asparagine
disulfide bond	A: 272-283	
mutagenesis site	A: 20	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	A: 116	W->A. Not expressed.
mutagenesis site	A: 187	W->A. Loss of ability to bind sphingomyelin, and induces hemolysis only at high concentration.
mutagenesis site	A: 206	W->A. Does not affect binding, and has little loss of hemolytic activity.
mutagenesis site	A: 245	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	A: 291	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
Site	C: 20	Crucial for binding sphingomyelin and inducing hemolysis
Site	C: 187	Crucial for binding sphingomyelin and important for inducing hemolysis
Site	C: 209	Important for activity
Site	C: 245	Crucial for binding sphingomyelin and inducing hemolysis
Site	C: 291	Crucial for binding sphingomyelin and inducing hemolysis
glycosylation site	C: 248	N-linked (GlcNAc...) asparagine
disulfide bond	C: 272-283	
mutagenesis site	C: 20	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	C: 116	W->A. Not expressed.
mutagenesis site	C: 187	W->A. Loss of ability to bind sphingomyelin, and induces hemolysis only at high concentration.
mutagenesis site	C: 206	W->A. Does not affect binding, and has little loss of hemolytic activity.
mutagenesis site	C: 245	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	C: 291	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
Site	B: 20	Crucial for binding sphingomyelin and inducing hemolysis
Site	B: 187	Crucial for binding sphingomyelin and important for inducing hemolysis
Site	B: 209	Important for activity
Site	B: 245	Crucial for binding sphingomyelin and inducing hemolysis
Site	B: 291	Crucial for binding sphingomyelin and inducing hemolysis
glycosylation site	B: 248	N-linked (GlcNAc...) asparagine
disulfide bond	B: 272-283	
mutagenesis site	B: 20	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	B: 116	W->A. Not expressed.
mutagenesis site	B: 187	W->A. Loss of ability to bind sphingomyelin, and induces hemolysis only at high concentration.
mutagenesis site	B: 206	W->A. Does not affect binding, and has little loss of hemolytic activity.
mutagenesis site	B: 245	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	B: 291	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
Site	D: 20	Crucial for binding sphingomyelin and inducing hemolysis
Site	D: 187	Crucial for binding sphingomyelin and important for inducing hemolysis
Site	D: 209	Important for activity
Site	D: 245	Crucial for binding sphingomyelin and inducing hemolysis

Site	D: 291	Crucial for binding sphingomyelin and inducing hemolysis
glycosylation site	D: 248	N-linked (GlcNAc...) asparagine
disulfide bond	D: 272-283	
mutagenesis site	D: 20	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	D: 116	W->A. Not expressed.
mutagenesis site	D: 187	W->A. Loss of ability to bind sphingomyelin, and induces hemolysis only at high concentration.
mutagenesis site	D: 206	W->A. Does not affect binding, and has little loss of hemolytic activity.
mutagenesis site	D: 245	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.
mutagenesis site	D: 291	W->A. Loss of ability to bind sphingomyelin, and to induce hemolysis.

Table 5: Interaction of residues of lysenin (3ZXD) with Shighella membrane protein 4Q35: A[%]

Residue	Closest	Distance	Specific Interactions	# HB	# Salt Bridges	# Pi Stacking	# Disulfides	# vdW Clash	Surface Complementarity
A:278:Tyr	Z:83:His	3.6 A		0	0	0	0	0	0.95
A:329:Val	Z:147:Gln	3.5 A		0	0	0	0	0	0.95
A:414:Ala	Z:57:Thr Z:58:Hip	3.5 A 3.8 A		0	0	0	0	0	0.94
Z:145:Ile	A:315:Trp A:313:Phe A:287:Met	3.3 A 3.6 A 3.7 A		0	0	0	0	0	0.93
A:315:Trp	Z:145:Ile Z:147:Gln	3.3 A 3.4 A		0	0	0	0	0	0.92
A:287:Met	Z:143:Gln Z:145:Ile	3.3 A 3.7 A		0	0	0	0	0	0.91
A:279:Leu	Z:83:His	3.2 A		0	0	0	0	0	0.9
Z:83:His	A:279:Leu A:255:Met A:278:Tyr A:254:Asn	3.2 A 3.5 A 3.6 A 3.7 A		0	0	0	0	0	0.9
A:394:Leu	Z:58:Hip Z:17:Val Z:121:Asp	3.4 A 3.7 A 3.9 A		0	0	0	0	0	0.89
Z:148:Val	A:327:Phe A:361:Val A:282:Ala	3.1 A 3.6 A 3.7 A		0	0	0	0	0	0.89
A:276:Phe	Z:142:Thr Z:143:Gln	3.4 A 3.9 A		0	0	0	0	0	0.88
A:309:Arg	Z:29:Arg	3.5 A		0	0	0	0	0	0.88
A:327:Phe	Z:148:Val Z:149:Ile Z:147:Gln	3.1 A 3.4 A 3.8 A		0	0	0	0	0	0.88
Z:29:Arg	A:291:Tyr A:309:Arg	3.0 A 3.5 A	1x hb to A:291:Tyr	1	0	0	0	0	0.88
Z:146:Pro				0	0	0	0	0	0.88
Z:58:Hip	A:394:Leu A:414:Ala	3.4 A 3.8 A		0	0	0	0	0	0.87
Z:139:Ile	A:291:Tyr	3.8 A		0	0	0	0	0	0.87
A:291:Tyr	Z:29:Arg Z:140:Gly Z:139:Ile	3.0 A 3.3 A 3.8 A	1x hb to Z:29:Arg	1	0	0	0	0	0.86
A:361:Val	Z:18:Ala Z:17:Val Z:148:Val Z:19:Val	3.2 A 3.3 A 3.6 A 3.9 A		0	0	0	0	0	0.86
Z:57:Thr	A:430:Val A:414:Ala A:413:Gln	3.4 A 3.5 A 3.8 A		0	0	0	0	0	0.86
Z:149:Ile	A:327:Phe A:321:Met A:325:Trp	3.4 A 3.5 A 4.0 A		0	0	0	0	0	0.86
Z:62:Ser	A:368:Phe A:398:Tyr A:367:Asn	3.1 A 3.3 A 3.9 A		0	0	0	0	0	0.84
A:255:Met	Z:83:His	3.5 A		0	0	0	0	0	0.83
A:317:Hie	Z:147:Gln	3.5 A		0	0	0	0	0	0.83
A:321:Met	Z:149:Ile	3.5 A		0	0	0	0	0	0.83
A:430:Val	Z:57:Thr	3.4 A		0	0	0	0	0	0.83
Z:147:Gln	A:315:Trp	3.4 A		0	0	0	0	0	0.83

	A:317:Hie A:329:Val A:327:Phe	3.5 A 3.5 A 3.8 A							
A:391:Glu	Z:21:Lys	3.1 A	2x hb, 1x salt bridge to Z:21:Lys	2	1	0	0	0	0.82
Z:150:Thr	A:325:Trp	3.3 A		0	0	0	0	0	0.82
Z:151:Ser	A:325:Trp	2.9 A		0	0	0	0	0	0.82
A:398:Tyr	Z:62:Ser	3.3 A		0	0	0	0	0	0.8
A:325:Trp	Z:151:Ser Z:150:Thr Z:15:Asp Z:152:Arg Z:16:Val Z:149:Ile	2.9 A 3.3 A 3.6 A 3.6 A 3.7 A 4.0 A		0	0	0	0	0	0.79
Z:21:Lys	A:391:Glu	3.1 A	2x hb, 1x salt bridge to A:391:Glu	2	1	0	0	0	0.79
A:396:Val	Z:59:Ser	3.6 A		0	0	0	0	0	0.78
Z:152:Arg	A:325:Trp	3.6 A		0	0	0	0	0	0.78
Z:19:Val	A:361:Val	3.9 A		0	0	0	0	0	0.76
Z:140:Gly	A:274:Asn A:291:Tyr A:289:Leu	3.2 A 3.3 A 3.8 A		0	0	0	0	0	0.76
A:282:Ala	Z:148:Val	3.7 A		0	0	0	0	0	0.75
A:372:Val	Z:17:Val Z:121:Asp	3.5 A 4.0 A	1x clash to Z:17:Val	0	0	0	0	1	0.75
A:281:Gln	Z:85:Glu	2.9 A	1x hb to Z:85:Glu	1	0	0	0	0	0.74
Z:17:Val	A:361:Val A:372:Val A:394:Leu	3.3 A 3.5 A 3.7 A	1x clash to A:372:Val	0	0	0	0	1	0.74
Z:15:Asp	A:363:Tyr A:325:Trp	3.3 A 3.6 A		0	0	0	0	0	0.73
Z:18:Ala	A:361:Val	3.2 A		0	0	0	0	0	0.73
A:311:Trp				0	0	0	0	0	0.72
Z:59:Ser	A:396:Val	3.6 A		0	0	0	0	0	0.72
Z:85:Glu	A:281:Gln	2.9 A	1x hb to A:281:Gln	1	0	0	0	0	0.72
A:363:Tyr	Z:15:Asp	3.3 A		0	0	0	0	0	0.71
A:368:Phe	Z:62:Ser Z:61:Gly	3.1 A 3.4 A		0	0	0	0	0	0.71
A:432:Leu	Z:66:Thr	3.5 A		0	0	0	0	0	0.7
Z:121:Asp	A:394:Leu A:372:Val	3.9 A 4.0 A		0	0	0	0	0	0.69
Z:142:Thr	A:276:Phe	3.4 A		0	0	0	0	0	0.69
A:254:Asn	Z:83:His	3.7 A		0	0	0	0	0	0.68
A:313:Phe	Z:145:Ile	3.6 A		0	0	0	0	0	0.68
Z:16:Val	A:325:Trp	3.7 A		0	0	0	0	0	0.65

[%]the interacting residues with a cut off value of surface complementarily selected at 0.65. Data obtained using Schrodinger protein-protein docking tool.

Fig 1A

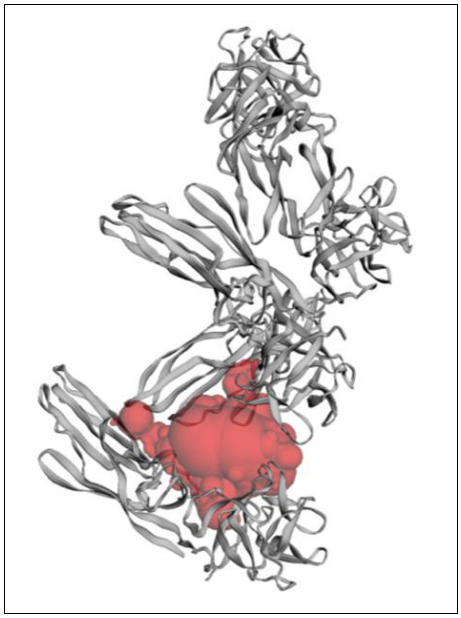

Fig 1B

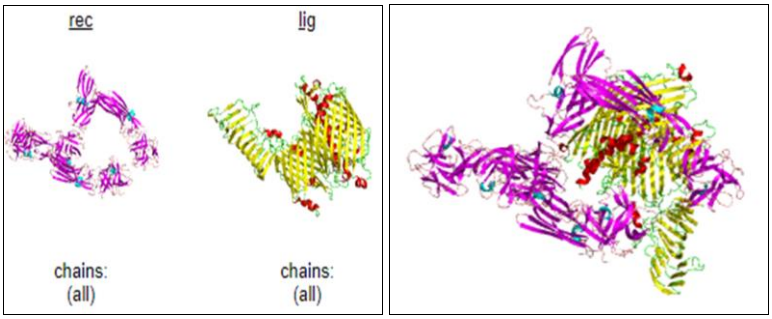

Fig 2

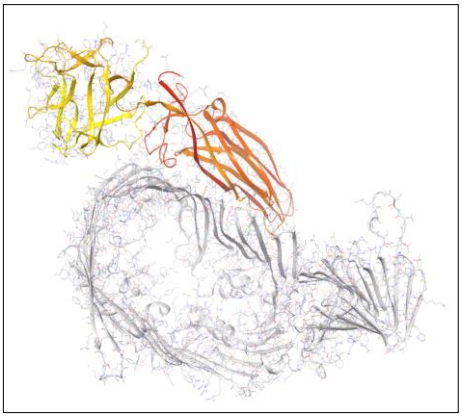

Fig 3A

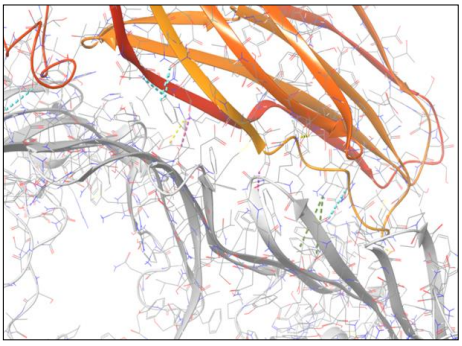

Fig 3B

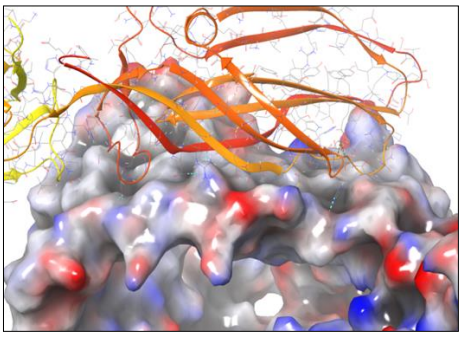

Fig 3C

Discussions

World Health Organisation (WHO) in 2017 (WHO/EMP/IAU/2017.12), has enlisted microbes for prioritization of research for new antibiotics due to rapid generation of antibiotic resistance including *Shighella* which reveal growing resistance to ampicillin, sulfonamides, nalidixic acid fluoroquinolone, and remains the major cause of morbidity and mortality globally, more-so recorded from developing countries with reports of 165 million cases globally, claiming an estimated around one million lives annually and is considered as a serious threat to human health by the Centre for Disease control and prevention (CDC).

The binding of lysenin wild type protein to *Shigella* membrane protein offers promises as antibacterial agents targeting *Shigella* pathogen. We have highlighted in this study the surface topology of monomeric unit of wild type lysenin and the different amino acids in the sequence and their roles and this study remains the first study in elucidating the binding of *Shigella* membrane protein with Lysenin by *insilico* and quantum statistical and mechanical approaches. Lysenin is known as a pore forming protein. The monomeric unit of lysenin was chosen to understand the complexity of interaction with the Shighella membrane protein in details. The importance of this study lies in highlighting the antimicrobial potential of this molecule. Since this was the first study we tried to understand the binding using different approaches of protein-protein docking using tools Cluspro, Schrodinger and in the light of quantum statistical and mechanical approaches and its limitations.

The predicted *O-ß*-GlcNAc attachment sites in the wild type lysenin including 32S, 54T, 77S, 110S, 143S, 149S, 149T, 150S, 182S, 190T, 218S, 225S, 237T with most predictability in sites 54T and 110S (Table 3) could play important role in the lysenin wild type-*Shigella* membrane protein interaction.

Using the *insilico* protein-protein docking study (Figures 2-3, Table 5) we have highlighted the binding of lysenin with *Shigella* and the interacting residues are detailed in Table 5 with a cut off value of surface complementarily selected at 0.65 and through the quantum mechanical understanding we have highlighted the probable implications of using lysenin as a molecule to target *Shigella*.

This study finds importance in the current times since there are no new antibiotics in the last few years and the alarming global concerns over increasing antibacterial resistance calls in for more research for new and effective antimicrobial targets**.**

It remains our future endeavour to find its application as drugs to *Shigella* infection.

Acknowledgements

The author acknowledges, NISER Bhubaneswar, India for the study. The author acknowledges the support of Evaluation liscence of Schondinger in the study.

References

1. Belotserkovsky I, Brunner K, Pinaud L, Rouvinski A, *et al*. Glycan-glycan interaction determines Shigella tropism toward human T lymphocytes. mBio, 2018, 9:e02309-17.
2. Bilej M, Procházková P, Šilerová M, *et al*. Earthworm Immunity. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience, 2000-2013.
3. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S, *et al*. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics, 2004; 4:1633-1649.
4. Chen VB, Arendall WB 3rd, Headd JJ, *et al*. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica, 2010, D66:12-21.
5. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978; 47:45-148.
6. Gasteiger E, Hoogland C, Gattiker A, *et al*, Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, 2005, 571-607.
7. Ghosh S. Insilico studies on Antimicrobial Peptides (AMPs) from Earthworm, Int J Pept Res Ther, 2019.
8. Ghosh S. Insilico study of earthworm CCF1 peptides in earthworm Int J Pept Res Ther, in press, 2020.
9. Ghosh S. Environmental pollutants, pathogens and immune system in earthworms, ESPR. 2018; 25:6196-6208.
10. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pacific Symposium on Biocomputing. 2002; 7:310-322.
11. https://www.expasy.org/proteomics
12. https://www.genome.jp/tools/motif/
13. https://www.schrodinger.com/maestro
14. https://www.schrodinger.com/protein-preparation-wizard
15. https://www.schrodinger.com/science-articles/biologics-design
16. Jennison AV, Verma NK. Shigella flexneri infection: pathogenesis and vaccine development. FEMS Microbiol Rev. 2004; 28:43-58.

17. Kobayashi H, Ohta N, Umeda M. Biology of lysenin, a protein in the coelomic fluid of the earthworm Eisenia foetida. Int Rev Cytol. 2004; 236:45-99.
18. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990; 276:172-174.
19. Kozakov D, Hall DR, Xia B, *et al*. The Clus Pro web server for protein-protein docking. Nature Protocols. 2017; 12:255-278.
Kozakov D, Beglov D, Bohnuud T, *et al*. How good is automated protein docking? Proteins: Structure, Function, and Bioinformatics. 2013; 81:2159-2166
20. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res 2:2. Jespersen MC, Peters B, Nielsen M, *et al*. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res (Web Server issue), 2006, 2:2.
21. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol. 1993; 234(4):946-950.
22. Monigatti F, Gasteiger E, Bairoch A, *et al*. The Sulfinator: predicting tyrosine sulfation sites in protein sequences Bioinformatics. 2002; 18:769-770.
23. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986; 25:5425-5432.
24. Prioritization of pathogens to guide discovery, Research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis, WHO/EMP/IAU/2017.12, https://www.who.int/medicines/areas/rational_use/PPLreport_2017_09_19.pdf?ua=1
25. Prochazkova P, Roubalova R, Skanta F, *et al*. Developmental and Immune Role of a Novel Multiple Cysteine Cluster TLR from Eisenia andrei Earthworms. Front. Immunol, 2019, 10:1277.
26. Qiao S, Luo Q, Zhao Y, *et al*. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. Nature. 2014; 511(7507):108-11.
27. Saha S, Raghava GPS. Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. Proteins. 2006; 65:40-48.
28. Swiderska B, Kedracka-Krok S, Panz T, *et al*, Lysenin family proteins in earthworm coelomocytes - Comparative approach. Dev Comp Immunol. 2017; 67:404-412.
29. Tian W, Chen C, Lei X, *et al*. CASTP 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res. 2018; 46(W1):W363-W367.