Comparative analysis of the molecular mechanism of heat shock proteins from pathogenic and non-pathogenic *E.coli*

Sanchari Bhattacharjee, Mohana Saha, Rakhi Dasgupta*, Angshuman Bagchi*

Department of Biochemistry and Biophysics,

University of Kalyani

Kalyani, Nadia

Telephone: +919051689541, +919051948843

E-mail: Sanchari Bhattacharjee: sancharib88@gmail.com

Mohona Saha: mohanasaha06@gmail.com

Rakhi Dasgupta: rdgadg@gmail.com

Angshuman Bagchi: angshumanb@gmail.com

***Corresponding authors**

Abstract:

Cells can withstand the effects of temperature stress by activating small heat shock proteins IbpA and IbpB. Lon protease employing Ser679 – Lys722 catalytic dyad proteolyze IbpA and IbpB in their free forms, at physiological temperature i.e. without any temperature stress. However, the proteolytic activity of IbpA and IbpB is affected when the catalytic dyad residue of Lon protease is mutated. The mutation S679A in Lon protease brings about some changes so that the proteolytic interactions between the small heat shock proteins with that of the mutant Lon protease are lost which makes a difference in the interaction pattern of mutant Lon protease with their substrates. In the present study, we made an attempt through in-silico approach to figure out the underlying aspects of the interactions between the small heat shock proteins IbpA and IbpB with mutant Lon protease in *Escherichia coli*. We have tried to decipher the molecular details of the mechanism of interaction of proteolytic machinery of small heat shock proteins and mutant Lon protease with S679A mutation at physiological temperature in absence cellular temperature stress. Our study may therefore be helpful to decode the mechanistic details of the correlation with IbpA, IbpB and S679A mutant Lon protease in *E. coli*.

Keywords: Heat Shock Response, IbpA, IbpB, Lon protease, Molecular Dynamics Simulations.

1. Introduction

Cells withstand different kinds of stresses while living in a variety of environments. These cellular stresses make the cellular proteins misfolded which affects proper cellular functioning. To maintain proper functioning, the cellular proteins must attain proper folding of misfolded proteins which is a grand challenge of nature. Cell with its own intrinsic anti-stress mechanism counteracts the stressed condition to ensure the proper folding. Cellular stresses are frequent and recurring which makes the cellular proteins to denature, aggregate and damage. The anti-stress response mechanism consists of different types of molecules which elevates at stressed cellular condition such as folding chaperones that ensure the native folding state of the misfolded proteins, Holdase and Disaggregases prevent the misfolding and disaggregation and Proteases eliminate the damaged proteins and thus play a pivotal role in maintaining cellular protein control [1].

Small heat -shock proteins (sHsps) are a type of chaperones with low molecular weight (12-30 kDa) so that are called as molecular chaperones. These proteins protect the stressed cellular protein from further misfolding or aggregation and then assign it to foldase chaperones to regain the native state of the misfolded protein. Two members of small heat shock family, IbpA and IbpB are initially considered as inclusion body protein but later it is observed that they expressed at heat-stressed condition to counter heat-stressed proteins from being misfolded i.e. they play their role as holdase. Both IbpA and IbpB show 48% similarity at amino acid level with a conserved -crystallin domain flanked N- and C- termini [2].

Lon being the ATP- dependent principal protease in *Escherichia coli* [3] possess a homohexameric structure with a protease and an ATPase domain. Homohexamers of Lon create an internal chamber which triggers degradation damaged proteins [4]. It is reported that at physiological temperature in absence of cellular stress IbpA and IbpB remain in their free form and serves as substrates of Lon protease which depict an interrelation between sHsps and protein degradation machinery [3]. It is observed that, Lon uses a unique Ser⁶⁷⁹ – Lys⁷²² catalytic dyad in active site within the protease domain to execute catalysis [5]. This catalytic pocket of Lon protease with the unique Ser-Lys dyad residues plays an important role in substrate binding as well as proteolytic degradation. It is observed and reported that, within the proteolytic domain, Lon protease has the classical catalytic triad residues H665, H667 and D676. But site directed mutagenesis of Lon protease containing substitutions in these classical catalytic triad residues i.e. D676N, H665Y, and H667Y impair the ATPase activity of the enzyme [6]. So the classical catalytic triad residues have no role in proteolysis. Surprisingly, it is also observed that within the proteolytic domain, Lon protease uses a unique conserved Ser679 – Lys722 dyad in active site to promote catalysis [5]. This catalytic pocket of the protease domain with Ser-Lys dyad plays a pivotal role in substrate binding. According to an experimental analysis, it is reported that if the catalytic dyad residues Lysine722 and Serine 679 are mutated to Glutamine and Alanine respectively by site directed mutagenesis, then these mutations affect proteolytic activity of Lon protease.

In our previous work, the mode of interactions between IbpA, IbpB and Lon protease is observed at physiological as well as cold and heat-shock temperature and it is observed that Lon interacts more positively with IbpAB at physiological temperature, than those of the stressed temperatures. In the present work, a comparison at physiological temperature has been made through *in-silico* approach to identify the pattern of interactions between IbpA and IbpB with wild type Lon protease as well as mutant Lon protease (S679A) in both pathogenic and non-pathogenic *E.coli* strains. to conclude with a plausible model of molecular mechanism between small heat shock proteins and protein degradation system by Lon protease. So this is the first report to compare and predict the mode of interaction schemes between these proteins in *Escherichia coli* through *in-silico* approach.

2. Materials and Methods

2.1 Sequence Homology Search and Modeling of the proteins: The amino acid sequences of IbpA, IbpB and Lon protease (UniProt Accession ID: P0C054, P0C058 ,P0A9M0 and A0A152UUM3 respectively) and their basic information such as sequence length, molecular mass, amino acid modifications were retrieved from UniProt (<http://www.uniprot.org/>) [7]. In case of UniProt ID P0A9M0, the expression organism of Lon protease is *Escherichia coli* K-12 strain which is a non pathogenic strain whereas in case of UniProt ID A0A152UUM3, the expression organism of Lon protease is *Escherichia coli* O157:H7 strain which is a pathogenic one.

The sequences were used to search for suitable templates in Protein Data Bank (PDB)(www.rcsb.org/) [8] using NCBI PSI-BLAST search tool [9]. For scoring parameters BLOSUM62 matrix is used with the gap cost Existence: 11 Extension: 1 and threshold value is 0.005. In case of both the IbpA and IbpB , the sequence similarity of the obtained PDB templates from BLAST search are less than 30 percent as there was no crystal structure of IbpA and IbpB of *Escherichia coli* in PDB. So, the FASTA sequences of IbpA and IbpB are

deposited in RaptorX (raptorx.uchicago.edu/) [10] modelling tools which generated automated models of the proteins. In case of IbpA, RaptorX selected the Heat shock protein 16.0 from *Schizosaccharomyces pombe* (PDB ID 3W1Z) [11] and for IbpB, RaptorX selected the crystal structure of eukaryotic small heat shock protein (PDB ID 1GME) [12] as best template from the database with the highest value of significance. These modelling servers are able to build the entire model of these proteins. The generated models of IbpA and IbpB by RaptorX show an uGDT score of 89 and 88 respectively, a p-value of 1.17e-06 and 1.44e-06 respectively with no disordered regions. The modelled protein with high p-value and uGDT score is likely to be a good quality model, so the parameter values of both the modelled proteins indicate that the generated models are with good qualities.

To authenticate the generated model secondary structure prediction is also done using PSIPRED secondary structure prediction server (<http://bioinf.cs.ucl.ac.uk/psipred/>). For the proteolytic activity, the α -crystallin domain residues are important because Lon protease recognizes this α -crystallin domain and promote proteolysis. It is predicted by PSIPRED Version 3.3 [13] that the α -crystallin domain of IbpA and IbpB spans from amino acid residue 35 to amino acid residue 120 with seven α -sheets and the generated model is also showed the similar secondary structure. So the generated model significantly matches with the predicted structure [Figure 1.A&B].

In case of Lon protease, both the FASTA sequences retrieved from non-pathogenic *E.coli* strain (UniProt Accession ID P0A9M0) and pathogenic *E.coli* strain (UniProt Accession ID A0A152UUM3) were used to search for suitable templates in Protein Data Bank (PDB) (www.rcsb.org/) [7] using NCBI PSI-BLAST search tool [8]. For scoring parameters, BLOSUM62 matrix was used with the gap cost: Existence: 11 Extension: 1 with a threshold value of 0.005. In case of both the sequences, the A chain of PDB ID 6U5Z [14] showed the highest query coverage and percentage identity. In case of Uniprot ID P0A9M0 i.e. Lon protease of non pathogenic *E.coli* strain, PDB ID 6U5Z would show 100 % query coverage, 99.87% percentage identity and an E value of 0.0 whereas in case of Uniprot ID A0A152UUM3 i.e. Lon protease of pathogenic *E.coli* strain, PDB ID 6U5Z would show 100 % query coverage, 99.74% percentage identity and an E value of 0.0. Homology modelling of these proteins were performed using SWISS-MODEL (<https://swissmodel.expasy.org/>) [15] homology modelling server using this PDB ID 6U5Z as a template to generate template based models of the proteins.

The stereo-chemical qualities and the conformations of the resultant structures of IbpA, IbpB and Lon protease were checked with PROCHECK [16] and Verify3D [17] through SAVES server at <http://nihserver.mbi.ucla.edu/SAVES/> and Ramachandran plots [18] were drawn. There were no residues present in the disallowed regions of Ramachandran plots. The Verify3D score of IbpA, IbpB and Lon protease would pass the value of 1D-3D profile.

2.2 Pairwise Sequence Alignment (PSA) of Lon protease of *E.coli* K-12 strain and Lon protease of *E.coli* O157:H7 strain

FASTA sequences of Lon proteases of *E.coli* K-12 and *E.coli* O157:H7 strains were submitted to Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [19] to perform Pairwise Sequence Alignment (PSA).

2.3 Site-directed mutagenesis and secondary structure superimposition of Lon protease: It is reported that if the catalytic dyad residue Serine679 is mutated to Alanine i.e. S679A, this mutation leads to the loss of proteolytic activity [20]. So, site-directed mutagenesis was performed in Discovery Studio version 2.5 to mutate these two Lon proteases at that particular position. The generated mutated structures were energy minimized and tested with PROCHECK and Verify3D. Then these wild type and mutated structures were superimposed using Superpose server (wishart.biology.ualberta.ca/SuperPose/) [21].

2.4 Molecular docking with IbpA, IbpB and Lon protease: The models of IbpA and IbpB were docked using Zdock docking server at (<http://zdock.umassmed.edu>) [22]. The best 10 docked complexes with best docking score were selected and checked with PROCHECK and Verify3D. The docked complex with lowest energy values was selected as final working model.

Next, the above mentioned docking procedures were again performed to dock the IbpAB docked complex with the crystal structure of wild type as well as mutant Lon protease separately in Z-Dock server. Again top 10 docked complexes with best docking scores are selected and in each of these docked complexes are minimized and the energy values are calculated. The docked complexes were tested using PROCHECK and Verify3D. The

docked complex with lowest energy scores is selected as final docked complex (IbpAB-Lon protease) [Table 1] [Table 2].

Table 1: The best 10 docking poses of IbpA, IbpB and wild type Lon protease showing all energy profiles, PROCHECK and Verify 3D score. The selected docked structure is marked in bold.

Docked complex	Interaction Energy (IE) (Kcal/Mol)	Van-der-Wal Energy (VDW) (Kcal/Mol)	Electrostatic Energy (EIE) (Kcal/Mol)	Binding Free Energy (BFE) (Kcal/Mol)	RAMPAGE Score (% of disallowed region)	Verify 3D Score
Complex 1	-163.36001	-23.92609	-139.43392	-158.1531	0%	89.83%
Complex 2	-157.59543	-25.32017	-132.27527	-152.8694	0%	82.74%
Complex 3	-146.95611	-20.41863	-126.53749	-145.135	0%	94.57%
Complex 4	-136.38073	-21.26018	-115.12054	-144.0759	0%	90.21%
Complex 5	-137.56429	-25.04836	-112.51593	-143.432	0%	90.98%
Complex 6	-131.06914	-22.30468	-108.76446	-144.6112	0%	82.62%
Complex 7	-152.20318	-23.26713	-128.93605	-154.5237	0%	87.92%
Complex 8	-146.59749	-17.5625	-129.03499	-147.2263	0%	88.78%
Complex 9	-161.39836	-22.13509	-139.26328	-153.1479	0%	96.90%
Complex10	-172.71306	-23.37996	-149.3331	-164.0855	0%	98.95%

Table 2: The best 10 docking poses of IbpA, IbpB and mutant Lon protease showing all energy profiles, PROCHECK and Verify 3D score. The selected docked structure is marked in bold.

Docked complex	Interaction Energy (IE) (Kcal/Mol)	Van-der-Wal Energy (VDW) (Kcal/Mol)	Electrostatic Energy (EIE) (Kcal/Mol)	Binding Free Energy (BFE) (Kcal/Mol)	RAMPAGE Score (% of disallowed region)	Verify 3D Score
Complex 1	-87.07811	-67.68525	-19.38986	-114.0491	0%	94.24%
Complex 2	-141.93129	-80.60985	-61.32144	-152.3293	0%	93.19%
Complex 3	1316.27711	1366.98106	-50.70395	1321.4310	0%	92.67%
Complex 4	-60.10451	-40.27554	-19.82897	-68.5817	0%	93.19%
Complex 5	-114.02465	-87.57776	-26.44689	-126.2604	0%	93.19%
Complex 6	19794.33187	19812.49892	-18.16706	19779.4255	0%	100%
Complex 7	3926.60519	3968.08510	-41.47991	3934.1060	0%	100%
Complex 8	-150.098	-111.811	-38.287	-113.5368	0%	92.67%
Complex 9	-190.37358	-73.69234	-116.68124	-215.9242	0%	94.24%
Complex10	-97.57960	69.80257	-167.38218	-85.3113	0%	92.15%

2.5 Molecular Dynamics Simulation of the IbpA-IbpB-Lon protease docked complex: In total, two systems (wild type docked complex and mutant type docked complex) were prepared for simulation. All MD simulations were performed in GROMACS 4.6.5 [23] applying Charmm27 force field [24]. Then, these structures were solvated by explicit solvent SPC/E system [25] water model in cubic boxes with a minimum 10Å distance from the edge of the cube. To obtain a neutral simulation cell the water molecules were replaced by the addition of counter ions (Na⁺ and Cl⁻). Each system was then minimized using steepest descent. The minimized docked complex was again tested with PROCHECK and Verify3D.

A 100 ps NVT equilibration was performed at 310 K with position restraints applied to all of the backbone atoms in order to relieve any bad contacts at the side chain solvent interface. Then position restraints were withdrawn and a 100 ps NPT simulation was conducted. Then 100ns production MD was performed. Snapshots of the trajectory were taken in every 2 ps.

In our previous study, it is reported that, Lon protease interacts with its substrates IbpA and IbpB more efficiently at physiological temperature (310K). In order to study the interaction pattern, both the docked complexes (i.e. wild type Lon -IbpAB and mutant Lon -IbpAB) were performed 100ns of production run in GROMACS version 4.6.5, keeping the pressure constant. The changes in conformation were recorded and plotted against time. Further analyses were done by VMD version 1.9.2 [26] and DS suite. MS Excel was used to generate the graph and for structure visualization, DS Visualizer was used. Relative solvent accessibility (RSA) values are calculated using the XSSP server (<http://www.cmbi.ru.nl/xssp/>) [27] and superimposition is done using the SuperPose server (<http://wishart.biology.ualberta.ca/SuperPose/>) [21]. Finally, these docked complexes were tested using PROCHECK and Verify3D.

2.6 System specification

All the work was done using windows operating system and the simulation was performed by Dell Workstation with 2x10 core xenon CPU, 32GB SCC RAM(4X8), 2X2 TB HDD, 1x5GB Nvidia Tesla K20C, 1X1 GB Nvidia Quadro K600.

3. Result and Discussions

3.1. Sequence similarity and RMSD changes in secondary structure

Pairwise Sequence Alignment reveals that amino acid sequences of Lon protease of *E.coli* non pathogenic K-12 and pathogenic O157:H7 strains are highly conserved [Supplementary file 1: Pairwise Sequence alignment of Lon protease of *E.coli* pathogenic O157:H7 strain and *E.coli* non pathogenic K-12 strain]. The Percentage Identity Matrix of these two sequences is as follows [Table 3]:-

Table 3: Percentage Identity Matrix of Lon protease of *E.coli* pathogenic O157:H7 strain and *E.coli* non pathogenic K-12 strain.

tr A0A152UUM3 A0A152UUM3_ECO57	100.00	99.87
sp P0A9M0 LON_ECOLI	99.87	100.00

Table 4: RMSD values of superimposed secondary structures of Wild type and S679A Mutant type Lon protease of Non-pathogenic (K-12 strain) and pathogenic (O157:H7 strain) variants of *E.coli*

	Wild type Lon (Pathogenic O157:H7 strain)	S679A Mutant type Lon (Pathogenic O157:H7 strain)
Wild type Lon (Non-pathogenic K-12 strain)	1.63 Å	3.21 Å
S679A Mutant type Lon (Non-pathogenic K-12 strain)	3.52 Å	1.58 Å

Pairwise sequence alignment shows that Lon protease is highly conserved whether it is present in pathogenic strain or in non pathogenic strain of *E.coli*. So, the secondary structures of Lon protease of the pathogenic variants and non-pathogenic variants would not have much significant differences as revealed by their backbone RMSD values whereas the Wild type and S679A mutant type Lon protease of the same strain can produce much more significant RMSD values [Table 4]. So, changes in the RMSD values clearly depicts that effect of S679A mutation in the catalytic pocket of Lon protease to have a significant effect on the overall RMSD value and causes more significant RMSD changes in secondary structure irrespective of pathogenicity or non-pathogenicity of the strain.

As the effect of S679A point mutation in the catalytic pocket of Lon protease imposes more significant effect in comparison to the pathogenicity or non-pathogenicity of the organism, so the remaining study is focussed on the wild type and S679A mutant type variants of *E.coli* K-12 strain.

3.2. Changes in catalytic pocket

In spite of having the catalytic triad residues (His665, His667 and Asp676) Lon protease does not form the catalytic triad rather it employs a unique Ser⁶⁷⁹- Lys⁷²² catalytic dyad for proteolysis. This catalytic dyad residues form a catalytic pocket with a unique structural fold to exert proteolysis. As one of the catalytic dyad residue is mutated (Ser 679Ala), a significant change is shown in the mutated catalytic pocket with respect to the wild type catalytic pocket.

The total SASA (Solvent Accessible Surface Area) value of the catalytic dyad of wild type Lon is 5840.7Å² whereas that of the mutant type catalytic dyad is 5912.3 Å². The SASA value is calculated to point out local structural perturbations due to this point mutation for each position of catalytic pocket before and after the mutation. RSA value analysis [Supplementary file 2: RSA value analysis of catalytic dyad] shows that there is a tendency of transition of core residues to surface residues in the catalytic pocket of mutant type Lon and due to this transition of solvent accessibility, the shape of the catalytic pocket is changed making it wider in the mutant Lon protease.

Superimposition of wild type catalytic pocket and mutant type catalytic pocket generates an overall RMSD change of 1.39Å which considered as a significant change. So, the SASA value, RSA values and RMSD values of these wild type and mutant type catalytic pocket depicts that the mutation in one of the catalytic dyad residue (Ser679Ala) of Lon forms a change in shape in the catalytic pocket of mutant Lon protease which may affect the substrate binding to catalytic pocket as well as the proteolytic activity [Figure 1.C-E].

Figure 1 Predicted secondary structure of [A].IbpA and [B].IbpB by PSIPRED version 3.3 (helix is marked as pink, strand as yellow, coil as black line; blue bars denote the confidence of prediction). - crystalline domain (from residue 35 to residue 120) of IbpA and IbpB are composed of seven -sheets (marked as S1 to S7). [C]

Catalytic pocket of wild type Lon protease (marked in red). [D] Catalytic pocket of mutant (S679A) Lon protease (marked in black). [E] Superimposition of catalytic groove of wild type and mutant type Lon protease.

3.3. Changes in binding pose

Binding site analysis showed a change in binding pose of wild type complex and mutant type complex [Figure 2.A & 2.B]. In wild type complex, Lon interacts with IbpB much more than IbpA whereas in case of mutant type complex, Lon interacts with IbpA and IbpB almost equally. Lon actually recognizes these two proteins equally but processes IbpB much faster than IbpA. The tails of IbpA inhibits Lon degradation, so Lon recognizes both these IbpA and IbpB equally but interacts and processes IbpB more as IbpB tails do not inhibit degradation. So, in wild type, Lon interacts with IbpB more and initiates proteolytic degradation. But in case of mutant type, as a result of catalytic dyad residue mutation (S679A), mutant Lon being unable to proteolyse affects the interaction of Lon protease with IbpA and IbpB.

Superimposition of wild type and mutant type docked complex generates an overall RMSD change of 2.97Å which considered as a significant RMSD change. So, these binding pose analysis and RMSD value calculation clearly depicts that the catalytic dyad residue mutation (S679A) of Lon protease affects the involvement of Lon protease with IbpA and IbpB and causes a binding pose alteration in mutant type complex in comparison to wild type complex [Figure 2.C].

3.4. Changes in pattern of interaction

3.4.1 Changes in the binding site

Lon protease interacts with IbpA & IbpB in their free form i.e. at physiological temperature whereas the N and C terminal residues of IbpA & IbpB are involved in oligomerization and chaperone activity in temperature stress. Lon recognizes the conserved α -crystallin domain of IbpA & IbpB to promote proteolysis. Binding site analysis revealed that in mutant type, more amino acid residues are involved in binding in corresponding to the wild type. Domain specific binding site analysis showed that in case of wild type complex, the α -crystallin domain residues are involved more in number than N and C terminal residues [Figure 2.D & 2.E] which clearly depicts that the wild type Lon protease initiates proteolysis of IbpA and IbpB at physiological temperature by involving α -crystallin domain residues. But in case of mutant type, it is observed that the N and C terminal residues of IbpA and IbpB are interacting more than α -crystallin domain residues. The mutation in one of the catalytic dyad residue (S679A) hampers the proteolytic activity of Lon protease so that Lon recognize less number of α -crystallin domain residues. In absence of such proteolytic degradation pathway, IbpA and IbpB interact with each other and oligomerize involving N and C terminal residues more in numbers.

3.4.2 Changes in the hydrogen bond interaction

The study reveals that there is a change in the pattern of hydrogen bond interactions. Hydrogen bond analysis shows that more numbers of hydrogen bonds are formed in mutant type in corresponding to the wild type [Figure 2.F]. Wild type Lon protease recognizes the α -crystallin domain of IbpA and IbpB to execute proteolysis. The processing of proteolytic degradation gives an unstable interaction with less numbers of hydrogen bonds. While in case of mutant type, due to the mutation (S679A) in catalytic dyad, the proteolytic activity is hampered partially and this makes mutant Lon unable to recognize the α -crystallin domain residues of IbpA & IbpB properly. Due to this improper recognition, mutant Lon cannot execute the proteolytic degradation properly. Probably, IbpA and IbpB in presence of such partial stress cannot degrade proteolytically but on the other hand interact with each other and form oligomers and makes a stable interaction with more numbers of hydrogen bonds.

Figure 2 Binding pose of [A] Wild type (WT) and [B] mutant (MT) complexess. Docked Ibpa and Ibpb are marked in black and Lon protease is in blue and green. Globular pink region indicates the binding surface. [C] Superimposition of WT and MT (IbpAB-Lon). Bar graphs showing [D] No. of amino acid residues involved in binding. [E] Domain specific binding residues of Ibpa & Ibpb (- crystalin domain [CD] and N & C terminal [N&C]). [F] No. of hydrogen bonds involved in WT & MT.

3.5. Changes in the energy profile

A comparison between the energy profiles of the wild type and mutant type complex was carried out. To analyze this, 148 frames of stable conformations along with the structures are retrieved from molecular dynamics trajectories for each of the docked complexes. Then of these 296 docking poses (148 docked complexes for wild type and mutant type each) the binding free energies (BFE), total interaction energies (TIE), Van-der-Waals energies (VDWE) and electrostatic interaction energies (EIE) are calculated in DS suite. The energy values are tabulated and the unpaired t-test is performed. The unpaired one tailed t-test shows that the alterations of these two data sets are significantly different at $P_{0.001}$ [Figure 3]. It is found that the values of energy profiles are significantly lower in mutant type in comparison to the wild type which states that the mutant type complexes have much more stable interactions throughout the molecular dynamics simulations [Table 5]

Table 5: Mean values (Kcal/mol) (mean $\pm$ SD) at p-value 0.001 and Sample size for all, n = 148 of the energy profiles of wild type and mutant type complexes.

type	BFE	TIE	VDWE	EIE
WT	-169.261 $\pm$ 58.43088	-137.177 $\pm$ 22.58188	-24.6837 $\pm$ 2.922644	-112.493 $\pm$ 23.45457
MUT	-401.29 $\pm$ 174.1517	-325.176 $\pm$ 41.70572	-56.1827 $\pm$ 9.832986	-268.994 $\pm$ 38.86405

Figure 3 Comparison of interaction energies (Kcal/Mol) (mean $\pm$ SD) of wild type (WT)(grey bar) and mutated (MUT)(black bar) complexes. [A] BFE [B] TIE [C] VDWE [D] EIE. An unpaired one tailed t-test showed all the mutated energy values are significantly less than wild type values ($P_{0.001}$).

3.6. Changes in secondary structure

There also a change is observed in secondary structure of wild type complex and mutant type complex. RMSF values of atoms of wild type and mutant type complexes showed that mean square fluctuation of atoms are much less in mutant type in comparison to the wild type. Less RMS fluctuation points a stable pattern of interaction. In mutant type, due to mutation in one of the catalytic dyad residue (S679A), Lon could not recognize the conserve α -crystallin domain residues of IbpA & IbpB properly and so unable to promote proteolytic degradation properly which in turn gives a stable interaction pattern with less RMS fluctuation.

Secondary structure analysis revealed that the mutant complex confers more structural stability in corresponding to the wild type complex. Wild type Lon with the help of the catalytic dyad residues exerts proteolytic degradation of IbpA and IbpB which give energetically unfavourable interaction with production of more numbers of coils. On the other hand, the mutant type Lon (S679A) due to it's mutation in catalytic dyad, cannot

degrade IbpA and IbpB proteolytically. In absence of this proteolytic degradation which ought to be exerted by Lon protease, IbpA and IbpB interact and oligomerize with each other forming an energetically favourable and comparatively stable structure with more numbers of helices, sheets and turns [Figure 4].

Figure 4 Scatter plot showing atom wise RMSF changes in [A] Wild type complex (red lines) and [B] Mutant type complex (blue lines). Pattern of changes in secondary structure of [C] wild type complex and [D] mutant type complex during simulation where Turn is marked as green, -sheet as yellow, Alpha helix as pink, 3_{10} helix as blue and Coil as white.

So, the mutant type complex generates more stable secondary structural pattern which clearly depicts the stable interaction pattern of IbpA and IbpB in absence of proteolytic degradation of mutant Lon protease.

So, the mutation in catalytic dyad residue (S679A) in Lon protease makes a difference in the relative solvent accessibility, total solvent accessible surface area that changes the shape of catalytic pocket which in turn affects the substrate binding to Lon protease for initiation of proteolysis. It also causes the alteration of binding pose in comparison to the wild type complex. This mutation also changes the pattern of interaction in mutant type making more stable and favourable interaction throughout the molecular simulation than the wild type which reflects in their hydrogen bond interaction, involvement of binding sites, interaction energy profiles as well as in the values of root mean square fluctuation. Moreover, the secondary structure analysis clearly points that energetically favourable structures are formed more in mutant Lon as there is loss of function due to this mutation. So, S679A mutation in Lon protease generates an overall change in structure and function which makes a change in the interaction pattern of Lon protease with their substrate IbpA and IbpB in *E. coli* even at physiological temperature.

4. Conclusion:

In this work, we tried to compare the difference in molecular mechanism of pathogenic and non-pathogenic *E.coli* strains. Previously, we made a comparison of the binding interactions in presence of double mutations, viz., S679A and K722Q [28]. However, in this case we did not find much difference between the pathogenic and non-pathogenic *E.coli* strains. Our study is the first of its kind which is based on direct analysis of the heat shock response in both the pathogenic and non-pathogenic *E.coli* strains.

5. Acknowledgement

Authors are thankful to the DBT Project No.BT/PR6869/BID/7/417/2012 for financial help and the generous help of Bioinformatics Infrastructural Facility (BIF), University of Kalyani, by providing the necessary machinery requirements and support.

6. Conflict of Interest

None

7. References

- [1] Jiao W, Hong W, Li P, Sun S, Ma J, Qian M, Hu M, Chang Z (2008) The dramatically increased chaperone activity of small heat-shock protein IbpB is retained for an extended period of time after the stress condition is removed. *Biochem. J* 410: 63–70. doi:10.1042/BJ20071120.
- [2] Stróżecka J, Chruściel E, Górna E, Szymańska A, Zitkiewicz S, Liberek K (2012) Importance of N- and C-terminal regions of IbpA, *Escherichia coli* small heat shock protein, for chaperone function and oligomerization. *J. Biol. Chem* 287: 2843–2853. doi:10.1074/jbc.M111.273847.
- [3] Bissonnette SA, Rivera-Rivera I, Sauer RT, Baker TA (2010) The IbpA and IbpB small heat-shock proteins are substrates of the AAA+ Lon protease. *Mol Microbiol*. 75: 1539–1549. doi:10.1111/j.1365-2958.2010.07070.x.
- [4] Vieux EF, Wohlever ML, Chen JZ, Sauer RT, Baker TA (2013) Distinct quaternary structures of the AAA+ Lon protease control substrate degradation. *Proc. Natl. Acad. Sci. U. S. A.* 1–7. doi:10.1073/pnas.1307066110.
- [5] Botos I, Melnikov EE, Cherry S, Tropea JE, Khalatova AG, Rasulova F, Dauter Z, Maurizi MR, Rotanova TV, Wlodawer A, Gustchina A (2004) The Catalytic Domain of *Escherichia coli* Lon Protease Has a Unique Fold and a Ser-Lys Dyad in the Active Site. *J. Biol. Chem.* 279 : 8140–8148. doi:10.1074/jbc.M312243200.
- [6] Starkova NN, Koroleva EP, Rumsh LD, Ginodman LM, Rotanova T V (1998) Mutations in the Proteolytic Domain of *Escherichia coli* Protease Lon Impair the ATPase Activity of the Enzyme. *FEBS Lett.* 422(2):218-20. doi: 10.1016/s0014-5793(98)00012-x
- [7] Leinonen R, Garcia Diez F, Binns D, Fleischmann W, Lopez R, Apweiler R (2004) UniProt archive. *Bioinformatics* 20 : 3236–3237. doi:10.1093/bioinformatics/bth191.
- [8] Berman HM (2008) The Protein Data Bank: a historical perspective. *Acta Crystallogr. A.* 64 :88–95. doi:10.1107/S0108767307035623
- [9] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10 : 421. doi:10.1186/1471-2105-10-421.
- [10] Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7 : 1511–1522. doi:10.1038/nprot.2012.085.
- [11] Hanazono Y, Takeda K, Oka T, Abe T, Tomonari T, Akiyama N, Aikawa Y, Yohda M, Miki K (2013) Nonequivalence observed for the 16-meric structure of a small heat shock protein, SpHsp16.0, from *Schizosaccharomyces pombe*. *Structure*. 21: 220–8. doi:10.1016/j.str.2012.11.015.
- [12] van Montfort RL, Basha E, Friedrich KL, Slingsby C, Vierling E (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* 8 :1025–1030. doi:10.1038/nsb722.

- [13] Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res.* 33 : W36-8. doi:10.1093/nar/gki410.
- [14] Botosa I, Lountos GT, Wud W, Cherry S, Ghirlando R, Kudzhayev AM, Rotanova TV, de Val N, Tropea JE, Gustchina A, Wlodawer A (2019) Cryo-EM structure of substrate-free E. coli Lon protease provides insights into the dynamics of Lon machinery. *Current Research in Structural Biology*.1:13-20. doi.org/10.1016/j.crstbi.2019.10.001.
- [15] Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31(13): 3381–3385. PMCID: PMC168927. PMID: 12824332. doi: 10.1093/nar/gkg520.
- [16] Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* 8 : 477–486. doi:10.1007/BF00228148.
- [17] Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature.* 356 : 83–85. doi:10.1038/356083a0.
- [18] Ramachandran GN, Ramakrishnan C, Sasisekharan V. (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol.* 7: 95–9. PMID: 13990617 DOI: 10.1016/s0022-2836(63)80023-6.
- [19] Sievers F, Higgins DG (2014) Clustal Omega .*Curr Protoc Bioinformatics.* 48:3.13.1-16. doi: 10.1002/0471250953.bi0313s48. PMID: 25501942.
- [20] Fischer H, Glockshuber R (1993) ATP hydrolysis is not stoichiometrically linked with proteolysis in the ATP-dependent protease La from *Escherichia coli*. *J. Biol. Chem.* 268 : 22502–7. <http://www.ncbi.nlm.nih.gov/pubmed/8226758>.
- [21] Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* 32: W590–W594. doi:10.1093/nar/gkh477.
- [22] Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z (2014) ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics.* 30 btu097. doi:10.1093/bioinformatics/btu097.
- [23] Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4 : 435–447. doi:10.1021/ct700301q.
- [24] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4 :187–217. doi:10.1002/jcc.540040211.
- [25] Lee MS, Salsbury FR, Olson MA (2004) An efficient hybrid explicit/implicit solvent method for biomolecular simulations. *J. Comput. Chem.* 25 : 1967–1978. doi:10.1002/jcc.20119.
- [26] Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* 14 : 33–38. doi:10.1016/0263-7855(96)00018-5.
- [27] Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22: 2577–637. doi:10.1002/bip.360221211.
- [28] Bhattacharjee S. , Halder, I, Mitra, S, Dasgupta, R, Bagchi, A (2019) Elucidation of the molecular mechanism of heat shock proteins and its correlation with the effects of double mutations S679A & K722Q in the catalytic dyad residues of Lon protease. *Gene Reports.* <https://doi.org/10.1016/j.genrep.2019.100438>