

RESEARCH ARTICLE

Structural, Functional and Phylogenetic Analysis of Sperm Lysozyme-Like Proteins

Shalini Kalra, Mangottil Ayyappan Pradeep[‡], Ashok K. Mohanty, Jai K. Kaushik*

BTIS Sub-DIC, Animal Biotechnology Centre, National Dairy Research Institute, Karnal, 132001, India

[‡] Current address: Marine Biotechnology Division, Central Marine Fisheries Research Institute, Kochi-682018, Kerala, India

* jaikr1@gmail.com



OPEN ACCESS

Citation: Kalra S, Pradeep MA, Mohanty AK, Kaushik JK (2016) Structural, Functional and Phylogenetic Analysis of Sperm Lysozyme-Like Proteins. PLoS ONE 11(11): e0166321. doi:10.1371/journal.pone.0166321

Editor: Alexander Wlodawer, NCI at Frederick, UNITED STATES

Received: July 16, 2016

Accepted: October 26, 2016

Published: November 10, 2016

Copyright: © 2016 Kalra et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The sequences were deposited in NCBI nucleotide database with accession numbers JN130361 (Lyzl2), JN130363 (Lyzl3/Spaca3), HQ285241 (Lyzl4), JN130364 (Lyzl5/Spaca5) and JN130362 (Lyzl6).

Funding: SK was supported by the Junior and Senior Research Fellowships of Council for Scientific and Industrial Research, New Delhi, for doctoral program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Sperm lysozyme-like proteins belonging to c-type lysozyme family evolved in multiple forms. Lysozyme-like proteins, viz., LYZL2, LYZL3 or SLLP1, LYZL4, LYZL5 and LYZL6 are expressed in the testis of mammals. Not all members of LYZL family have been uniformly and unambiguously identified in the genome and proteome of mammals. Some studies suggested a role of SLLP1 and LYZL4 in fertilization; however, the function of other LYZL proteins is unknown. We identified all known forms of LYZL proteins in buffalo sperm by LC-MS/MS. Cloning and sequence analysis of the *Lyz*/cDNA showed 38–50% identity at amino acid level among the buffalo LYZL paralogs, complete conservation of eight cysteines and other signature sequences of c-type lysozyme family. Catalytic residues in SLLP1, LYZL4 and LYZL5 have undergone replacement. The substrate binding residues showed significant variation in LYZL proteins. Residues at sites 62, 101, 114 in LYZL4; 101 in SLLP1; 37, 62, and 101 in LYZL6 were more variable among diverse species. Sites 63 and 108 occupied by tryptophan were least tolerant to variation. Site 37 also showed lower tolerance to substitution in SLLP1, LYZL4 and LYZL5, but more variable in non-testicular lysozymes. Models of LYZL proteins were created by homology modeling and the substrate binding pockets were analyzed in term of binding energies and contacting residues of LYZL proteins with tri-N-acetylglucosamine (NAG)₃ in the A-B-C and B-C-D binding mode. Except LYZL6, LYZL proteins did not show significant difference in binding energies in comparison to hen egg white lysozyme in the A-B-C mode. (NAG)₃ binding energy in the B-C-D mode was higher by 1.3–2.2 kcal/mol than in A-B-C mode. Structural analysis indicated that (NAG)₃ was involved in making more extensive interactions including hydrogen bonding with LYZL proteins in B-C-D mode than in A-B-C mode. Despite large sequence divergence among themselves and with respect to c-type lysozymes, substrate binding residues as well as hydrogen bonding network between (NAG)₃ and proteins were mostly conserved. LYZL5 in buffalo and other mammalian species contained additional 10–12 amino acid sequence at c-terminal that matched with ankyrin repeat domain-containing protein 27. Phylogenetic analysis indicated LYZL2 to be most ancient among all the LYZL proteins and that the evolution of LYZL proteins occurred through several gene duplications preceding the speciation of mammals from other vertebrates as distant as reptiles and amphibians.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Lysozymes are hydrolytic enzymes, which cleave the β -(1,4)-glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine (NAG) residues of peptidoglycan of bacterial cell wall. It can also cleave β -(1,4) glycosidic bond of polymer of NAGs in chitin which is found in cell wall of fungi, exoskeletons of crustaceans, arachnids and insects [1,2]. On the basis of biological origin, lysozymes are divided into several families: chicken (c-type), goose (g-type), invertebrate (i-type), bacteriophage and plant lysozymes. Among these, c-type lysozymes are the most common and widely distributed among insects, fishes, reptiles, birds and mammals. Lysozyme is secreted in many body fluids such as saliva, blood, tears and milk in mammals. It is present at very high concentration in the eggs of many bird species [3]. Lysozyme expression has also been observed in many other tissues, viz., stomach, kidney, lacrimal gland, parotid gland, sublingual gland, lung, spleen, small intestine, lymph node, placenta and leukocytes [4].

Lysozyme has evolved in different tissues to perform specific functions, e.g., it has evolved as a digestive enzyme in ruminants [5,6], leaf eating monkeys [7] and hoatzin bird [8]. The stomach lysozyme has evolved to resist the acidic environment and pepsin in stomach. Ruminant artiodactyls like cattle and sheep have approximately ten lysozyme genes, whereas non-ruminant artiodactyls like pig have only single gene [9–11]. The c-type lysozyme family also includes α -lactalbumin and calcium-binding lysozyme. α -lactalbumin shares 40% sequence similarity with lysozyme but lacks bacteriolytic activity [12]. It is restricted to mammals and expressed specifically in lactating mammary glands to help in the synthesis of lactose [13]. Bacteriolytic calcium-binding lysozymes are found in the milk of few mammalian species (e.g. in dog, cat, seal, horse and echidna) and in the eggs (e.g. pigeon) and stomach of some leaf eating bird species like hoatzin [14]. A pseudogene *LYSC1* with calcium-binding property has also been found in some mammalian species [15]. It has been thought that calcium-binding lysozymes evolved from lysozyme before the divergence of birds and mammals; however, α -lactalbumin evolved only in mammalian clade.

Recently testis specific lysozyme-like genes (*Lyzl*) belonging to the c-type lysozyme family have also been discovered. Five forms of lysozyme-like genes *Lyzl2*, *Lyzl3/Spaca3*, *Lyzl4*, *Lyzl5/Spaca5* and *Lyzl6* have been identified in the testis of human and other mammalian species [16–19]. The lysozyme-like proteins (LYZL) along with commonly known lysozyme have been identified from sperm proteome of mouse and human [20,21]. LYZL4 protein has also been observed on the surface of human embryonic stem cells [22]. LYZL5/SPACA5 protein has been associated with Type-II diabetes as well [23]. Lysozyme, LYZL4, LYZL3/SPACA3 and LYZL5/SPACA5 have also been identified in the human sperm tail proteome [24]. LYZL3/SPACA3, LYZL4 and LYZL5/SPACA5 have been suggested to serve as biomarkers for male fertility [25–29]. LYZL3/SPACA3 protein is also known as sperm lysozyme-like protein 1 (SLLP1).

Lysozyme-like genes possess genomic organization and conserved signature sequences that are common among lysozyme family members [15,17]. In several LYZL proteins, the catalytic residues common among active c-type lysozymes are not conserved [16,18,30]. The role of multiple *Lyzl* genes in the male reproductive system is still obscure. The first LYZL protein SLLP1 encoded by *Spaca3* gene was reported to have a role in sperm-egg binding in human [16] and mouse [31]. Recently, presence of SLLP1 was observed in bovine oocytes and implicated in female fertility as well [32]. Mouse LYZL4 was also reported to be involved in male fertility [30]. LYZL1/LYZL2 and LYZL4 protein expression was found downregulated in asthenozoospermic condition in human [33]. Expression of several *Lyzl* genes has been observed in tissues like brain, lung, heart, spleen, kidney, ovary and uterus [17–19]; however their tissue-specific role is almost unknown. Functional characterization of LYZL proteins has

been restricted mainly to human and mouse in case of SLLP1 [16,31] and mouse in case of LYZL4 [30]. The functions of LYZL2, LYZL5 and LYZL6 are yet to be uncovered.

The evolution of numerous *Lyzl* genes and their functional role in mammals is still not well understood. Not all *Lyzl* genes are invariably distributed in all mammals and other vertebrates. In the present investigation, we identified all the known forms of lysozyme-like proteins in buffalo sperm lysate. The identified *Lyzl* genes were cloned, sequenced and comparative sequence analysis carried out to understand the effect of amino acid replacements on LYZL function. Further, structural models were built for LYZL proteins complexed with lysozyme inhibitor tri-N-acetylglucosamine (NAG)₃ to understand the structural and functional properties of the lysozyme-like proteins.

Materials and Methods

Isolation of total sperm proteins

Fresh semen sample of riverine buffalo (*Bubalus bubalis* Murrah breed) was obtained in sterile tubes from the semen collection service of the Artificial Breeding Research Centre, National Dairy Research Institute. The bull semen is regularly collected at the centre for breeding program or for discarding to maintain good semen quality in the bulls and hence no ethical clearance was required for use of semen sample. The semen was centrifuged at 9500×g for 10 min at 4°C. The supernatant was discarded and the pellet with sperm was washed with PBS buffer at pH 7.4 several times and mixed with radio immunoprecipitation assay lysis buffer (Cat. No. R0278, Sigma Chem. Co., USA) and incubated on ice for 45 min [34]. Thereafter, the mixture was centrifuged at 16000×g for 30 min at 4°C and the supernatant fraction was run on 12% SDS-PAGE and the gel was stained with 0.1% Coomassie brilliant blue G-250 in 30% methanol and 10% acetic acid.

Mass spectrometry analysis

The gel slices corresponding to sperm proteins of 12–18 kDa molecular weight were excised from acrylamide gel. Trypsin digestion of the extracted proteins was performed as described for in-gel digestion [35]. The peptide mixture was fractionated on a liquid chromatography (LC) system equipped with nano-trap column (C18, 5 μm, 200 Å, 0.1 x 150 mm) and a reverse phase nano-column (C18, 3 μm, 200 Å, 0.1 x 150 mm) in series. The mobile phase included 0.1% formic acid in water and 99.9% acetonitrile with 0.1% formic acid. A linear gradient of acetonitrile was set from 5% to 45% in 30 min at a flow rate of 800 nl/min. The peptides were submitted to the inline nanospray equipped ESI-tandem mass spectrometry (MS/MS) maXis HD ESI-QUAD-TOF (Bruker Daltonics, Germany). The mass spectrometer was configured for data-dependent acquisition mode for automatic selection of precursor ions for further analysis by MS in the m/z range of 50–2200. The generated MS spectra were used for searching against NCBI database by using Mascot software (Matrix Science). The peptide mass tolerance and MS/MS tolerance were set to 50 ppm and 0.05 dalton, respectively in the Mascot search.

Cloning of buffalo *Lyzl* genes

The primer pairs for the amplification of lysozyme-like genes were designed from cattle gene sequences with NCBI accession numbers BC126794 (*Lyzl2*), BC149686 (*Lyzl3/Spaca3*), BC111339 (*Lyzl4*), BC108157 (*Lyzl5/Spaca5*) and BC114039 (*Lyzl6*). Primers were designed from the regions outside the open reading frames (ORF) by using Primer3 software.

Buffalo testicular tissue was collected from the adult males slaughtered for the meat purpose at the Delhi slaughter house, Ghazipur, New Delhi. Animals were not slaughtered specifically

for tissue collection and hence ethical clearance was not required to collect the samples. Total RNA was extracted from the testicular tissue by using Trizol reagent (Invitrogen, USA) as per manufacturer's instructions. Phusion RT-PCR kit (Finnzymes, Finland) was used for the synthesis of first strand cDNA. Full strand cDNA was amplified by using gene specific primers and the PCR product was ligated with pJET1.2/blunt vector (Fermentas, USA) followed by transformation into chemically competent TOP10 cells (Invitrogen, USA). At least two positive clones confirmed by gene-specific PCR were used for sequencing. The recombinant DNA related procedures were approved by the local institute biosafety committee.

Protein sequence analysis

The ORF in various *Lyzl* genes were determined by using the NCBI ORF finder tool (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) followed by prediction of amino acid sequence using ExPASy translate tool (<http://www.expasy.org>). The signal peptide was predicted by using online SignalP 4.0 server (<http://www.cbs.dtu.dk/services/SignalP/>) based on neural network trained on eukaryotes [36]. The amino acids composition, molecular weight and isoelectric point (*pI*) of matured lysozyme-like proteins were predicted by using ExPASy ProtParam tool (<http://expasy.org/protparam/>). Signature sequence pattern of a lysozyme-like protein was identified with PROSITE database at ExPASy [37].

The multiple sequence alignment of buffalo LYZL proteins was performed by using the ClustalW2 program (www.ebi.ac.uk/Tools/msa/clustalw2) and pairwise sequence identities were determined between the orthologs and paralogs. The protein sequences of LYZL proteins of several species were retrieved from UniProtKB database (ExPASy) for sequence analysis. Non-testicular lysozymes included hen egg white lysozyme (HEWL), turkey egg white lysozyme (TEWL), hotzain bird, rainbow trout lysozyme (RBTL), Japanese quail, solea fish lysozyme and lysozymes from milk, stomach, mammary gland and also from other tissues of several species. Testicular lysozyme-like proteins included LYZL2, SLLP1, LYZL4, LYZL5 and LYZL6 from human, chimpanzee, gorilla, gibbon, baboon, monkey, orangutan, garnett's greater bushbaby, mouse, rat, mole rat, buffalo, cattle, sheep, goat, cat, dog, ground squirrel, rabbit, guinea pig, giant panda, mustela furo, african elephant, horse, camel, little brown bat, black flying fox, chinese hamster, american pika, armadillo, and european shrew (S1 Table). Protein sequences were analyzed for the identification of residues involved in catalysis and substrate binding. The substrate binding residues were analyzed among non-testicular lysozymes and testicular lysozyme-like proteins from several species.

For phylogenetic analysis, sequences were aligned with muscle program by using default parameters [38]. The phylogenetic tree of lysozyme-like proteins was generated with MEGA6 software [39] by using maximum likelihood method (ML) based on LG model [40] and by using a bootstrap value of 1000. Initial tree was obtained by heuristic search using Neighbor-Joining method as implemented in MEGA6 program. Five discrete gamma distribution categories were taken into account for evolutionary rates among sites.

Structural modeling of LYZL-(NAG)₃ complexes and estimation of binding energies

Homology models of LYZL proteins complexed with (NAG)₃ were built with YASARA program (version 16.4.8) [41] by using the protocol similar to that used earlier [42]. (NAG)₃ has been reported to bind in A-B-C or B-C-D subsites of c-type lysozymes, therefore, LYZL complex models with (NAG)₃ were created in both binding modes. The following template structures, HEWL (PDB IDs: 1LZB [43], 1HEW [44], 3A3Q [45]), TEWL (PDB ID: 1JEF) [46], human lysozyme (PDB ID: 1BB5) (unpublished) containing (NAG)₃ in the A-B-C binding

mode were used for LYZL-(NAG)₃ in the A-B-C mode. PDB ID 1BB5 contained two molecules and hence both were used in making the models after separating them. For B-C-D mode, complex structures with well-defined positions of (NAG)₃ were not available for HEWL, TEWL or human lysozyme. Superimposition of RBTL-(NAG)₄ complex structure in A-B-C-D binding mode (PDB ID: 1LMQ) [47] with RBTL-(NAG)₃ complex in the B-C-D mode (PDB ID: 1LMP) [47] showed NAG units in the B-C subsites to be completely overlapping [48] with minor change in NAG conformation at subsite D [47]. Similar observation has been made in HEWL complex, where B-C subunits of (NAG)₄ completely superimposed with that of (NAG)₃ in the partially occupied A-B-C and B-C-D binding modes [43], and with minor flexibility in position of NAG at subsite A. Therefore, we used HEWL (PDB ID: 1LZC) [43], (PDB ID: 1LSZ) [49], (PDB ID: 3WVY) [50], human lysozyme (PDB ID: 1LZR) [51] and RBTL (PDB ID: 1LMQ) [47] with (NAG)₄ in the A-B-C-D subsites and deleted NAG in subsite A to produce lysozyme-(NAG)₃ in B-C-D subsites. The RBTL complex structure with B-C-D binding mode (PDB ID: 1LMP) was also used as a template without any change.

The sequences of LYZL proteins along with the above complex templates were manually provided and alignment profiles were built by running PSI-BLAST [52]. Single model was created if alignment was certain, otherwise alternative models were built when alignment was ambiguous with a given template sequence. The missing loops were built and the side-chain rotamers optimized by taking into account electrostatic interactions, knowledge-based packing interactions and the solvation effect [53]. Initially, energy minimization was carried out with combined steepest descent and simulated annealing by fixing the backbone atoms of the aligned model. Thereafter, minimization with full unrestrained all-atom simulated annealing was carried out. The protocol used implicit solvent model for the optimization of side-chains and loops and explicit solvent shell model for the simulated annealing to fine tune the models. The overall quality Z-score, which is defined as the weighted averages of the individual Z-scores, was calculated by using the equation, $Z\text{-score} = 0.1456 \times \text{Dihedrals} + 0.3906 \times \text{Packing1D} + 0.4656 \times \text{Packing3D}$ [53]. The models with lower quality Z-score were discarded. The detailed stereochemical quality checks of the models were also carried out by using Procheck program.

The complex structures of LYZL with (NAG)₃ were superimposed and aligned by using MUSTANG program implemented in the YASARA program [54] and the RMSD of the models was calculated with respect to template structure. The selected models based on structural alignment and quality Z-score were further refined by all-atom MD using Yasara2 force field [53] with explicit solvent model. The periodic cell was filled with a water density of 0.997 and pK_a of titrable side chains was predicted [55] and assigned to respective groups before running MD. The MD was run at 298 K by controlling temperature with Berenden-type thermostat as implemented in the YARASA program where temperature of the solute and water were controlled independently to achieve uniform temperature in the cell. The trajectories were calculated with a time step of 2 fs and snapshots were saved at every 25 ps. The MD for refinement of all the initial homology models was run for 250 ps. For calculating binding energies of (NAG)₃ with LYZL proteins, all the 10 snapshots obtained during refinement for a given template were used to avoid bias toward specific conformation. Finally, binding energies of all the conformations generated from various templates were averaged out to avoid conformational and template biasness. Before calculating the binding energies of (NAG)₃ with LYZL proteins, hydrogen bonding network was optimized for all the LYZL-(NAG)₃ complex structures by using the Yasara method [56]. Binding energy and contacting residues of templates and LYZL-(NAG)₃ complexes were calculated by using molecular docking program VinaDock [57] with default parameters. The MD and docking protocol were set up in YASARA program [54].

Results

Identification of LYZL proteins in buffalo partial sperm proteome

LYZL proteins are constituted of 126–130 amino acid residues with molecular mass of 14–15 kDa. To identify lysozyme-like proteins in buffalo sperm, we analyzed sperm proteome subset corresponding to 12–18 kDa molecular mass by using LC-MS/MS. All LYZL proteins namely LYZL2, SLLP1, LYZL4, LYZL5 and LYZL6 could be identified in Mascot search.

Cloning and sequence analysis of *Lyzl* genes

The PCR amplified products of 519 bp (*Lyzl2*), 557 bp (*Lyzl3/Spaca3*), 532 bp (*Lyzl4*), 532 bp (*Lyzl5/Spaca5*) and 547 bp (*Lyzl6*) cloned into blunt end PCR cloning vector were sequenced. The sequences were deposited in NCBI nucleotide database with accession numbers JN130361 (*Lyzl2*), JN130363 (*Lyzl3/Spaca3*), HQ285241 (*Lyzl4*), JN130364 (*Lyzl5/Spaca5*) and JN130362 (*Lyzl6*). The open reading frame sizes of these genes were determined as follow, 447 bp (*Lyzl2*), 492 bp (*Lyzl3/Spaca3*), 438 bp (*Lyzl4*), 471 bp (*Lyzl5/Spaca5*) and 447 bp (*Lyzl6*). Precursors of LYZL2 and LYZL6 proteins consisted of N-terminal signal peptide of 19 amino acids and mature protein of 129 amino acids. SLLP1, which is a product of *Lyzl3/Spaca3* gene, possessed a signal peptide of 35 amino acids and mature protein of 128 amino acids. LYZL4 protein consisted of 19 residues in signal peptide and 126 residues in mature protein. LYZL5 consisted of a signal peptide of 18 amino acids and mature protein of 138 amino acids. SLLP1, LYZL5 and LYZL6 were observed acidic in nature with isoelectric points (pI) of 5.22, 5.55 and 5.15, respectively, whereas LYZL2 and LYZL4 were basic with pI of 8.13 and 8.46, respectively.

Pairwise sequence comparison showed that various buffalo LYZL proteins shared only 38–50% sequence identity with each other and other non-testicular c-type lysozymes like HEWL, human lysozyme, buffalo milk lysozyme and buffalo stomach lysozyme. Among several species, e.g., in human, chimpanzee, mouse, rat, cattle, buffalo, sheep and goat, orthologs shared high similarity. Identity of buffalo LYZL proteins was 96–100% with orthologs of ruminants, 75–88% with primate orthologs and 73–80% with rodent orthologs. Buffalo LYZL proteins showed highest sequence identity with cattle orthologs. SLLP1 and LYZL4 showed 100% identity, while LYZL5 showed lowest identity at 97% with the respective cattle proteins. LYZL2 and LYZL6 showed 99% identity with cattle orthologs. In most of the cases, buffalo LYZL2 showed highest identity, while LYZL5 showed lowest identity with their orthologs in primates and rodents.

Multiple sequence alignment of mature LYZL proteins showed that the eight cysteine residues forming four disulfide bonds in c-type lysozyme were conserved in all the buffalo LYZL proteins (Fig 1). LYZL5 contained an additional tail of 10 amino acids at the c-terminus (residues 129–138) as compared to other LYZL proteins. A deletion of five amino acids (residues 69–73) was observed in buffalo LYZL4. Similar deletion has also been observed in human, rat and mouse LYZL4 [18,30]. HEWL and most other non-testicular lysozymes contained six tryptophan residues that are conserved in only LYZL5, other LYZL proteins possessed only five conserved tryptophan residues. Some of the tryptophan residues are part of a hydrophobic box in c-type lysozymes. In HEWL, the box is made by Leu17, Tyr20, Tyr23, Trp28 and Trp111. All the buffalo LYZL proteins contained Leu17, Trp28 and Trp111. The residue numbering is as per HEWL, but due to addition and deletions, Trp111 has different residue number in LYZL proteins. Tyr20 was conserved in LYZL2 and SLLP1, but replaced with aromatic residue Phe in LYZL4, LYZL5 and LYZL6. Tyr23 was conserved in all LYZL proteins except in LYZL2, in which Phe23 was present. HEWL Trp62 is also conserved in turkey, buffalo, cattle, goat and sheep non-testicular lysozymes, while the non-testicular lysozymes in human,

chinese hamster, american pika, armadillo, mole rat and european shrew) possessed threonine (Thr35). A deviation in garnett's greater bushbaby and platypus was seen, they contained Ser35 and Glu35, respectively in SLLP1. In LYZL4 Glu35 was conserved in all mammals except in platypus that has alanine, while second catalytic residue Asp showed Asp53→Gly replacement in buffalo and other mammalian species. On the other hand, LYZL4 in platypus and armadillo possessed serine (Ser53). In LYZL5 both 35 and 52 positions were occupied by Glu in all the species except little brown bat which possessed valine (Val35) and glutamine (Gln35). In case of LYZL6, both the catalytic residues (Glu35 and Asp52) were conserved in buffalo and other species, except in horse which has glutamine at position 35.

Substrate binding pocket in LYZL proteins

The substrate binding pocket of lysozyme consists of residues including Asn37, Trp62, Trp63, Asp101, Trp108, and Arg114 in HEWL [2,16,31]. Three tryptophan namely Trp62, Trp63 and Trp108 play an important role in substrate binding [58,59]. Analysis of buffalo LYZL proteins showed that Trp63 and Trp108 were conserved in all the LYZL proteins as well; however, variations were observed in other substrate binding residues (Table 1). Trp63/108 were conserved in all the members of c-type lysozyme family as well, while variations have been observed at position 62 (Trp62) and other substrate binding residues [1,60]. Only LYZL5 possessed Trp62, which was replaced with aromatic Tyr in LYZL6, polar Thr in LYZL2, ionic Lys in SLLP1 and Asp in LYZL4. Polar Asn37 in HEWL was replaced with ionic Asp in LYZL6, polar His in LYZL2, and Gly in SLLP1 and LYZL5. Asp101 was replaced with Glu in LYZL2 and SLLP1, Ser in LYZL5, Gly in LYZL4 and LYZL6. Ionic Arg114 was replaced either with polar Asn in LYZL2 and LYZL4 or His in SLLP1, LYZL5 and LYZL6.

Table 1. Substrate binding residues in c-type lysozyme family.

Residue Number (according to HEWL seq)	Non-Testicular c-type Lysozymes	Testicular Lysozyme-Like Proteins from several species						
		LYZL2	SLLP1	SLLP1 (Reptiles)	LYZL4	LYZL5	LYZL5 (Frog)	LYZL6
37	N*/D/S/G/R (Cattle-intestine)	G*/H/R	G	G (Alligator, Anole)/H (Python, Garter snake)/R (Turtle)	K*/H (Dog)	G	R	K*/D/E (Pig)/H/N/R (Dog)/Y
62	W*/Y	T*/A/I (Platypus)/V (Orangutan)	K*/R/L (Platypus)	L*/I (Python)/Q (Turtle)	D*/E/A (Pika)/G (Bushbaby, Rabbit)/K (Guinea pig)/L (Squirrel)/V (Horse, Gibbon)	W	W	Y*/F (Pika, Rabbit)/V (Platypus)/L (European shrew)
63	W	W	W	W	W*/R (Molerat)	W*/C (Mole rat)	W	W
101	D*/G (Turkey)	E*/D (Pika, Rabbit)	E*/D/G/Q/R (Bushbaby)/V (Platypus)	G*/Q (Turtle, Alligator)	G*/D (Cat) /R (Flying fox)/E (Horse, Platypus)/T (Bushbaby)	S*/A (Squirrel)/L (European shrew, Hamster)/T (Rabbit)/Q (Platypus)	D	G*/A (Guinea pig)/E (Platypus)/R/S (Horse)
108	W	W	W*/R (Platypus)	W	W*/R (Dog)	W	W	W
114	H*/K/R	H*/N/R (Molerat)/Y (Elephant)	H*/N (Platypus)/	N*/H (Turtle, Alligator)	N*/F (Bushbaby)/H (Mole rat, Guinea pig)/Y	H*/N (Platypus)	Y	H*/N (Camel)/Y (Marmoset)

* The most occurring residues in species at respective positions, while those occurring in one or two species only are shown with species name in parenthesis.

doi:10.1371/journal.pone.0166321.t001

Analysis of the substrate binding residues of LYZL proteins across vertebrate species showed large variation at several positions (Table 1), although few positions were much more intolerant to variation than others. Residues at sites 62, 101, 114 in LYZL4; 101 in SLLP1; 37, 62, and 101 in LYZL6 were more variable. Sites 63 and 108, both occupied by tryptophan, were least tolerant to variation. Position 62 represented by tryptophan in most cases or tyrosine in few of the non-testicular lysozymes was conserved in LYZL5 of most species. This position was represented by threonine or alanine in LYZL2, lysine or arginine in SLLP1, aspartate or glutamate in LYZL4, and tyrosine in LYZL6 of most of species. Trp63 was almost completely conserved in all orthologs of all the LYZL proteins except in case of LYZL4 and LYZL5 of molarat. Site 37 also showed lower tolerance to changes in SLLP1, LYZL4, and LYZL5, although this site showed higher variability in non-testicular lysozymes.

The replacement of residues involved in binding with residues of significantly different physicochemical properties suggested that either the chemical nature of these residues might not be very crucial for binding of substrates or they have evolved for different subset of receptors/ligands.

LYZL proteins possess structural scaffold similar to c-type lysozymes

Comparative models of various LYZL proteins complexed with (NAG)₃ possessed structural topology similar to lysozyme (Fig 2). Mouse SLLP1 (mSLLP1) crystal structure (2GOI) solved at 2.30 Å has also been reported to contain glycan binding groove similar to lysozyme [61]. Buffalo SLLP1 shared 73% sequence identity with mSLLP1. Structural model of buffalo SLLP1 created with mSLLP1 template differed from SLLP1-(NAG)₃ complex based on six other lysozyme structures with RMSDs of 0.9±0.05Å. The observed RMSD between models created by using different template structures with considerably different sequences suggested the reliability of homology modeling protocol.

The average RMSD between models of buffalo LYZL proteins and the respective template structures was 0.5–0.8 Å. The major difference in the structures of various LYZL proteins was observed in the loop spanning residues 65 to 75. The conformation of this loop was also variable in template structures. This loop in buffalo LYZL4 was shorter by 5 residues. On the other hand, an insertion of one residue (Asp) was observed at position 50 in the β-turn connecting two β-strands in LYZL4 (residues number 51–55). In all other LYZL proteins, the loop size in β-turn was conserved. Another significant feature was observed in LYZL5 protein, which contained 10 extra amino acid residues at the c-terminus in comparison to other LYZL proteins. No function has been assigned to the segment; however, the conservation of the segment in LYZL5 of all the species suggested its functional implication. The BLAST search of the segment sequence in NCBI database revealed 90% identity with ankyrin repeat domain-containing protein 27, also known as Vps9 and ankyrin repeat containing protein (Varp) [62]. Ankyrin repeat-containing proteins constitute a large family of proteins involved in cellular interaction, communication, adhesion, signaling, vesicle fusion and molecular trafficking [62,63].

Analysis of substrate binding pocket of LYZL proteins

Fig 2 shows (NAG)₃ complexed with LYZL proteins in the A-B-C mode in the substrate binding pocket of LYZL proteins, similar to those present in template structures [43]. The possibility of the binding of (NAG)₃ in the B-C-D mode was explored if (NAG)₃ shifts position from A-B-C to B-C-D; however, no shift was observed in the molecular dynamics trajectory of 5000 ps simulation in explicit solvent. The existence of the B-C-D binding mode could not be rejected given the short simulation time. Long simulation considerably changed the conformation of (NAG)₃ and LYZL proteins. Therefore, this approach to explore existence of alternative

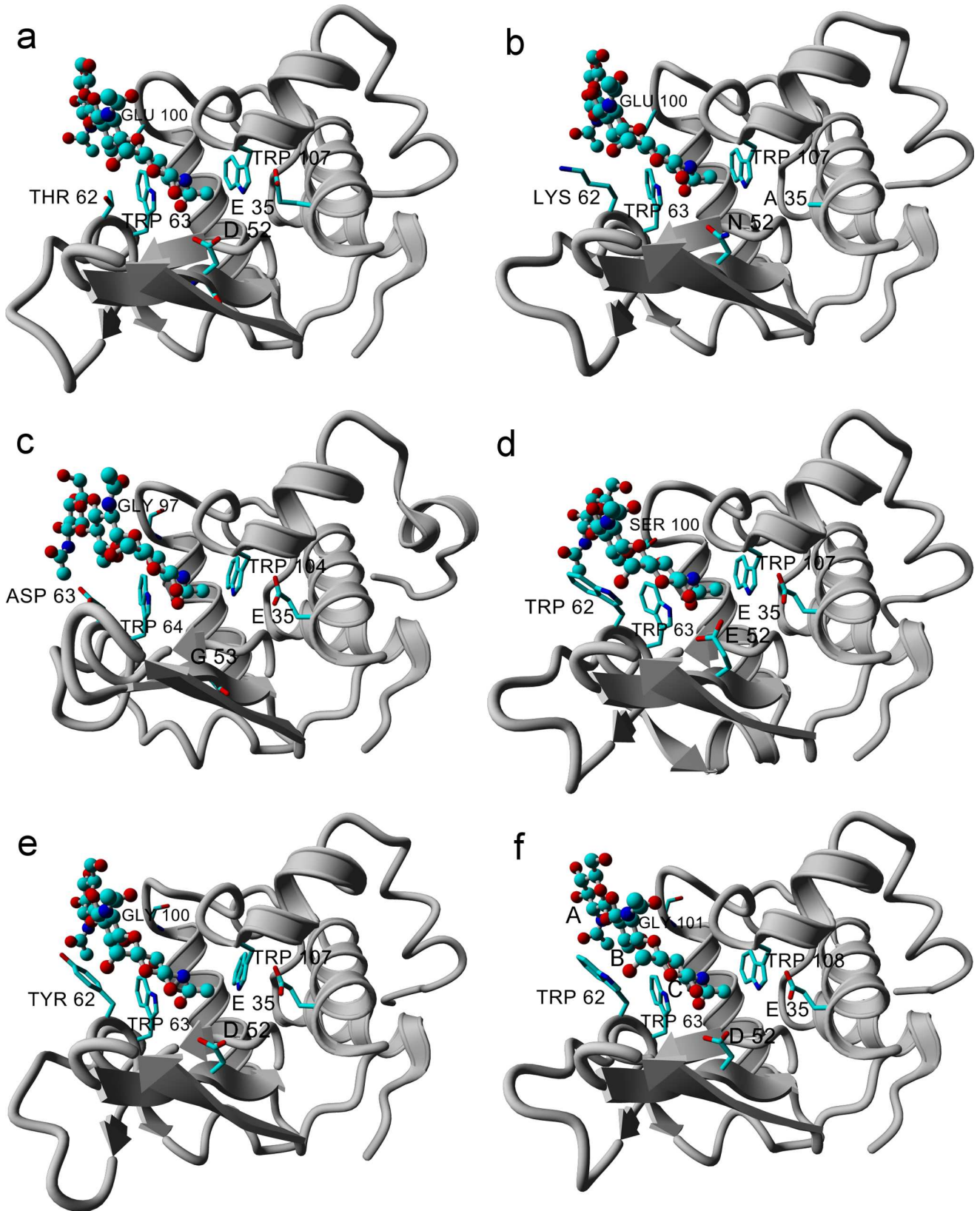


Fig 2. Structural models of lysozyme-like proteins complexed with (NAG)₃ based on several template structures. PDB IDs of template are provided in material method part. For clarity only one model for each LYZL protein is shown. Panels a–LYZL2, b–SLLP1, c–LYZL4, d–LYZL5, e–LYZL6 and f–1JEF (TEWL as one of the template). The protein part is shown in gray color and (NAG)₃ molecule in ball and stick style has been shown in elemental colors. The substrate binding residues interacting with (NAG)₃ are shown in three letter amino acid codes, while catalytic residues (residues at position 35 and 52 or 53) are labelled with single letter code. The NAG monomer binding subsites are represented by capital letters A, B and C in panel f.

doi:10.1371/journal.pone.0166321.g002

binding mode was dispensed with and homology models were created based on template with (NAG)₃ in the B-C-D binding mode and their binding energies were compared with those in the A-B-C mode.

In our calculations, the template complex showed an average binding energy of -5.7 kcal/mol for binding in the A-B-C subsites in the lysozyme-(NAG)₃ complex in the template structures. The result represents template complex structures of wild or mutants of human lysozyme, HEWL [43,44,45] and TEWL [46]. In some cases, the mutations were present at substrate binding site [45]. Moreover, human lysozyme possessed Tyr62 and TEWL possessed Gly101 in place of Trp62 and Asp101 in HEWL. Both these residues are known to stabilize substrate binding [64]. Asp101 is considered important for stabilizing the substrate binding since the Asp101→Gly replacement has been shown to decrease binding energy in HEWL [2]. The presence of Gly101 in TEWL resulted in the lowest binding energy with (NAG)₃ among all the template structures. This caused a marginally lower average binding energy of all templates together in comparison to HEWL. The human lysozyme has Tyr63 at position corresponding to Trp62 in HEWL, and the two molecules of human lysozyme 1BB5A and 1BB5B showed almost similar binding energy (-5.79 kcal/mol) with (NAG)₃. In case of HEWL, the binding energies were similar for complex templates 1LZB (-6.02 kcal/mol) and 3A3Q (-5.87 kcal/mol) in the A-B-C binding mode. Our results suggested that the obtained values was consistent with the relative affinities of (NAG)₃ for various lysozymes and analysis could be extended to LYZL proteins.

The experimental binding energy of (NAG)₃ with HEWL has been reported to be -7.0 kcal/mol [65], which is higher than the calculated value of -5.95 kcal/mol (average of 1LZB and 3A3Q PDBs templates). Nonetheless, these empirical values should provide relative binding strength of (NAG)₃ with various LYZL proteins. Table 2 lists the binding energies and contacting residues of LYZL proteins with (NAG)₃. The binding energies of LYZL2, SLLP1 and LYZL4 were insignificantly smaller by ≤ 0.5 kcal/mol in comparison to HEWL in the A-B-C mode. In case of LYZL6, the binding energies with (NAG)₃ in the A-B-C subsites was moderately lower (≥ 1.0 kcal/mol) than that of HEWL.

On the other hand, the binding energies of (NAG)₃ was always higher in the B-C-D binding mode than in the A-B-C mode for each of the LYZL proteins. The binding in the B-C-D mode was favored by 1.3 kcal/mol to 2.2 kcal/mol in comparison to binding in the A-B-C mode (Table 2). These differences were significant and suggested that B-C-D binding mode should be more favorable for binding of (NAG)₃ with LYZL proteins. Similar trend was also observed in case of template structures. Lysozyme templates showed a difference of > 1 kcal/mol between the two modes. Several of the templates were created by deleting NAG residue at the subsites A or D from a lysozyme-(NAG)₄ structure to create two binding modes and thus could have led to artefacts. To check it, we also calculated binding energy of HEWL-(NAG)₃ created by deleting NAG at site D of template 1LZC and observed that binding energy was similar to that calculated for X-ray structure of HEW-(NAG)₃ (1LZB). Similarly, binding energy of (NAG)₃ in B-C-D subsites created from (NAG)₄ complexed with RBTL (1LMQ) was similar to that obtained for X-ray structure of RBTL with (NAG)₃ in B-C-D mode (1LMP). Consistency of these results suggested that the artificially created complexes of (NAG)₃ with lysozymes did not introduce any artefact in the evaluated binding energies.

Table 2. Binding energy, contacting residues and hydrogen bonding residues of LYZL with (NAG)₃ in their complex structures.

LYZL-(NAG) ₃ complex	Binding Energy (kcal/mol) ±sd ^a	Contacting residues (ABC mode) ^b	Residues ^c involved in hydrogen bonding with (NAG) ₃
LYZL2-(NAG) ₃	5.50 ± 0.42	Asp52, Gln57, Phe56, Ile58, Asn59, Thr62, Trp63, Lys72, Arg74, Ile97, Glu100, Thr101, Asp102, Tyr106, Trp107, Gln108	Asn59, Thr62(8), Trp63, Glu100, Thr101(15), Asp102(8), Tyr106, Gln108(10)
SLLP1-(NAG) ₃	5.52 ± 0.45	Gln57, Ile58, Asn59, Arg61, Lys62, Trp63, Leu74, Ile97, Glu100, Pro101, Gln102, Ser106, Trp107	Asn59, Arg61(11), Trp63, Glu100, Gln102, Ser106
LYZL4-(NAG) ₃	5.43 ± 0.60	Phe57, Gln58, Ile59, Arg60, His62, Asp63, Trp64, Arg69, Arg71, Ile94, Gly97, Lys98, Arg99, Ala103, Trp104, Pro105	Arg60, Trp64, Arg71(14), Gly97(16), Arg99, Ala103
LYZL5-(NAG) ₃	5.32 ± 0.41	Glu52, Phe56, Gln57, Leu58, Asn59, Trp62, Trp63, Val97, Ser100, Glu101, Ser102, Ala106, Trp107, Asp108	Asn59, Trp62, Trp63, Glu101(25), Ala106, Asp108(15)
LYZL6-(NAG) ₃	4.83 ± 0.44	Asp52, Phe56, Gln57, Ile58, Asn59, Tyr62, Trp63, Asn74, Ile97, Gly100, Ala101, Gly102, Asn106, Trp107, Val108	Asn59, Trp63, Gly100(10), Gly102(8), Asn106
Template (1LZB)	6.02 ± 0.25	Asn46, Asp52, Gln57, Ile58, Asn59, Arg61, Trp62, Trp63, Arg73, Leu75, Ile98, Asp101, Gly102, Asn103, Ala107, Trp108	Asn59, Trp62, Trp63, Asp101, Asn103, Ala107
(BCD mode)^b			
LYZL2-(NAG) ₃	6.84 ± 0.39	Glu35, Leu46, Asp52, Gln57, Ile58, Asn59, Thr62, Trp63, Arg74, Ile97, Glu100, Tyr106, Trp107, Gln108, Gly109	Glu35, Asp52, Gln57(10), Asn59, Thr62(11), Trp63, Glu100, Tyr106, Gln108
SLLP1-(NAG) ₃	7.34 ± 0.47	Glu46, Asn52, Gln57, Ile58, Asn59, Arg61, Lys62, Trp63, Leu74, Ile97, Glu100, Gln102, Ser106, Trp107, Glu108, Ala109	Glu46, Asn52(21), Gln57(17), Asn59, Arg61(13), Trp63, Glu100(18), Gln102, Ser106, Glu108
LYZL4-(NAG) ₃	7.62 ± 0.47	Glu35, Tyr44, Asn46, Arg48, Phe57, Gln58, Ile59, Arg60, His62, Asp63, Trp64, Arg71, Ile94, Arg99, Ala103, Trp104, Pro105, Ser106	Glu35, Asn46(25), Gln58, Arg60, Asp63(21), Trp64, Arg71(13), Arg99, Ala103
LYZL5-(NAG) ₃	6.93 ± 0.33	Glu35, Asn44, Asn46, Glu52, Phe56, Gln57, Leu58, Asn59, Trp62, Trp63, Val97, Ser100, Ser102, Ala106, Trp107, Asp108, Ser109	Glu35, Glu52, Asn59, Trp62(16), Trp63, Ser102(8), Ala106, Asp108
LYZL6-(NAG) ₃	6.69 ± 0.41	Glu35, Asn44, Asn46, Asp52, Phe56, Gln57, Ile58, Asn59, Tyr62, Trp63, Asn74, Ile97, Asn106, Trp107, Val108, Lys109	Glu35, Asn46(17), Asp52, Gln57(9), Asn59, Tyr62(6), Trp63, Asn106, Val108
Template (1LMP)	7.0 ± 0.34	Glu35, Asn44, Asn46, Asp52, Gln57, Ile58, Asn59, Tyr62, Trp63, Lys73, Val75, Val98, Asp101, Asn103, Ala107, Trp108, Val109, Ala110	Glu35, Asp52, Asn59, Tyr62, Trp63, Asp101, Asn103, Ala107, Val109

^a sd—standard deviation.

^b The listed residues were observed interacting with (NAG)₃ in majority of the model structures. The residue numbering differ among the LYZL proteins due to insertions or deletions, see sequence alignment in Fig 1 for corresponding residues in various LYZL proteins. The data are based on analysis of 60 structures in each case of LYZL proteins in each of the binding mode.

^c Residues shown are those involved in making hydrogen bond with (NAG)₃ in most of the model structures. The numbers in parenthesis indicate the number of models in which a particular residue was observed in only a few models making hydrogen bond with (NAG)₃ out of total 60 models.

doi:10.1371/journal.pone.0166321.t002

Phylogenetic analysis of LYZL proteins

Gene tree was created with protein sequences by using ML method. Several substitution model for amino acid substitution were applied. Among them LG+G5+I evolutionary model provided the best likelihood values. Most of the important nodes showed reliable bootstrap values of >70% except for the bifurcation of branch leading to SLLP1 and LYZL5 subtrees (Fig 3). The gene tree suggested the duplication of ancestral c-type lysozyme gene into testicular and non-testicular lysozymes, the later one bifurcated further in to stomach and milk lysozymes in ruminants. On the other branch, multiple duplications and the divergence of the testicular genes led to evolution of multiple forms of lysozyme-like genes. First duplication in the ancestral *Lyzl* gene led to *Lyzl2* gene on one branch and ancestral gene of others on another clade. Subsequent duplication of the later ancestral gene resulted into *Sllp1* and *Lyzl5* on one branch and *Lyzl4* and *Lyzl6* on another. The gene duplication event leading to evolution of *Lyzl4* and *Lyzl6* was reliable; however, the ancestral node of *Sllp1* and *Lyzl5* showed lower bootstrap

Fig 3. Phylogenetic map of lysozyme family. The phylogenetic tree was constructed by using ML method. The nodes with diamond (◆) in blue color represent duplication event. LYZ-lysozyme, LYZL-lysozyme-like proteins, BoSt-cattle stomach, BuSt-buffalo stomach, ShSt-sheep stomach, BoMlk-cattle milk, BuMlk-buffalo milk, Hu-human, Ch-chimpanzee, Mc-monkey, Rt-rat, Mu-mouse, Bo-cattle, Bu-buffalo, Sh-sheep, Gt-goat, Py-python, Gr-garter snake. Lysozymes from non-mammals such as *Drosophila* (Dr) and *Bombyx mori* (Bm) were used as outgroups.

doi:10.1371/journal.pone.0166321.g003

value. Bayesian phylogenetic inference of the gene tree also suggested similar relationship of *Lyzl* genes (tree not shown) as shown by ML method. Gene tree clearly suggested that the duplication events leading to various LYZL paralog genes took place prior to speciation of mammalian species from reptiles and amphibians.

Discussion

Buffalo sperm possess all the known forms of LYZL proteins

All the known forms of LYZL proteins, namely LYZL2, LYZL3 or SLLP1, LYZL4, LYZL5 and LYZL6, were discovered in the buffalo sperm sub-proteome corresponding to 12–18 kDa molecular mass. Analyses of proteome of several species suggested that not all LYZL paralogs were present in sperm. However, this could also be because of inherent problems associated with efficiency of protein extraction, detection or identification associated with MS workflow. LYZL3 also named as SLLP1 has been detected in the sperm proteome of human [66–68], cattle [69,70] and horse [71]. LYZL4 has been detected in the sperm proteome of human [72,73], cattle [69,74] and boar [29]. The LYZL5/SPACA5 protein has been detected on human sperm [23,72]. Lysozyme, LYZL4, LYZL3/SPACA3 and LYZL5/SPACA5 were also identified in human sperm tail proteome [24].

Few of the sequenced genomes, especially those of mammals, suggested the presence of *Lyzl* genes. *Lyzl1* and *Lyzl2* are duplicated forms of the same gene and have been identified in only primates like human, chimpanzee, macaca, orangutan and common marmoset [15], while baboon possessed only a single gene. Both LYZL1 and LYZL2 have also been identified in human sperm proteome [21], which suggested that these genes could be functional copies. In northern treeshrew one copy of *Lyzl1* has been observed as functional, while the other one as a pseudogene [15]. *Spaca5* has also undergone duplication in human, while most primates and other mammalian species contain only a single copy. In buffalo sperm proteome we could identify only one form of each of LYZL2 and LYZL5. In the extreme case, genome of grey short tailed opossum, a marsupial, has been reported to have undergone deletion of *Lyzl1/2*, *Lyzl3/Spaca3*, *Lyzl6* genes, whereas *Lyzl4* and *Lyzl5/Spaca5* genes could not be identified in their genome [15]. We cloned and sequenced the cDNA of all the five buffalo *Lyzl* genes, which should be functional since their translated products were also detected in the buffalo sperm proteome. We could also detect all these proteins in the goat sperm proteome as well (unpublished data). Cattle also possessed all *Lyzl* genes as revealed by the presence of cDNA sequences or genomic sequences in the NCBI nucleotide database. These data suggest that there should be a strong evolutionary pressure for retaining *Lyzl* genes in mammals and several other vertebrates at high level of sequence conservation.

Bacteriolytic activity in LYZL proteins

Lysozyme has evolved in different tissues to perform a specific function. LYZL proteins evolved in testis in multiple forms. Mammary gland lysozymes are basic proteins having pI close to 9.0, whereas stomach lysozymes are acidic in nature. These proteins evolved and adapted to particular environment of the tissue. In case of testicular lysozymes, it is noteworthy

that LYZL2 and LYZL4 are basic proteins, while others are acidic proteins. The presence of both acidic and basic LYZL proteins in the male reproductive system is not clear given the identical environment in which they have evolved. LYZL2 (basic) and LYZL6 (acidic) possessed intact catalytic residues and could be possessing bacteriolytic activity, which has been shown at least in case of human LYZL6 [19]. The other three LYZL proteins are likely to be inactive because of mutation of catalytic residues.

Human SLLP1 did not exhibit catalytic activity because of Glu35→Thr and Asp52→Asn mutations [16]. Buffalo SLLP1 has also undergone changes at both of these positions (Glu35→Ala and Asp52→Asn), which might render it catalytically inactive, that was also indicated by our preliminary finding on recombinant buffalo SLLP1 expressed in yeast (unpublished results). Lack of bacteriolytic activity has also been observed in case of rat and mouse LYZL4 proteins that could also be inferred from the fact that both Glu35 and Asp52 are essentially required for catalytic activity [18,30]. LYZL5 contained Glu at both 35 and 52 positions. Analysis of lysozymes (Asp52→Glu) mutant (Glu35/Glu52) showed <3% bacteriolytic activity as compared to its wild type [75,76], which implies that buffalo LYZL5 might also possess very low bacteriolytic activity, if any.

Substrate binding pocket in LYZL proteins

Structural modeling of LYZL proteins with (NAG)₃ showed that these proteins carry substrate binding cleft similar to that present in active c-type lysozymes. Computational analysis of (NAG)₃ binding in the A-B-C subsites of various LYZL proteins showed binding energies of LYZL2 and SLLP1 almost similar to that of HEWL. LYZL4 and LYZL5 showed binding energy only moderately smaller by >0.5 kcal/mol and LYZL6 smaller by >1.0 kcal/mol in comparison to HEWL.

Asp101→Gly replacement in HEWL has been reported to change the conformation of the main-chain that forms a loop structure, which is part of the active site around subsite A. The loss of hydrogen bonding between (NAG)₃ and Asp101 in Asp101→Gly HEWL mutant has been observed to reduce the stacking interaction of Trp62 with (NAG)₃ [63]. Double mutants Trp62→His/Asp101→Gly [60] and Trp62→Tyr/Asp101→Gly [2] reduced the substrate binding affinity significantly. TEWL possesses Gly101 and the loop conformation has been observed similar to that seen in Asp101→Gly mutant of HEWL [46,64]. LYZL4 and LYZL6 have undergone Asp101→Gly replacement, while LYZL5 contains serine at corresponding position (Fig 2). These replacements are likely to affect the binding energy of (NAG)₃ with LYZL proteins at subsite A in the A-B-C mode. Except LYZL5, all the LYZL proteins have undergone replacement at position 62 (Table 1), which could also affect their affinities for (NAG)₃ in the A-B-C subsites. Trp62 mutant of HEWL showed enhanced rate of catalysis as well as dissociation constant [2,77–80]. Although Trp62 is important for cushioning and stabilizing substrate molecule, yet position 62 seems comparatively more tolerant to changes. We observed that position 62 is occupied by residues with dissimilar physicochemical properties in lysozymes of various species (Table 1). LYZL2 and SLLP1 possessing Glu100 showed higher affinity for (NAG)₃, which suggested that the presence of Asp or Glu at this position is important to stabilize binding of (NAG)₃ with LYZL proteins. Aspartate at this position could make up to two hydrogen bonds with (NAG)₃ to stabilize the protein-carbohydrate complex. LYZL6 was observed to make least number of hydrogen bonds with (NAG)₃ and also showed lowest binding energy (Table 2).

All the buffalo LYZL proteins possessed a conserved Trp at position corresponding to Trp63 of HEWL. This residue also makes extensive interaction with (NAG)₃ and replacement of Trp62 and/or Asp101, as in case of LYZL proteins, is likely to change the interaction of

Trp63 with (NAG)₃ as well. Given that Trp63 is almost 100% conserved in all LYZL proteins across the species except in case of molarat (Table 1), the residues corresponding to HEWL at position 62 and 101 should be deciding factor for the stability of complex. Previous studies suggested that even a single or double mutation in substrate binding pocket of lysozyme could affect the binding affinity of lysozyme for their substrates to varying extent [2,43,60, 64,81,82], while still maintaining the overall structure almost identical to wild type [51,64].

Binding energies of (NAG)₃ with LYZL proteins in the B-C-D mode were always higher than in A-B-C mode. Our results implied that binding of (NAG)₃ with LYZL proteins in B-C-D mode could be energetically more favorable, which is consistent with HEWL-(NAG)₃ showing higher binding energy in B-C-D mode [83]. These results implied that a dynamic equilibrium between population of A-B-C and B-C-D subsites occupied LYZL proteins might be existing as observed in the crystal structures of Trp62 and Asp101 mutants of HEWL [43]. However, the X-ray crystallographic studies showed wild type and some mutants of HEWL, TEWL and human lysozymes to contain (NAG)₃ in the A-B-C subsites [43–46]. This could be because of constraints placed by crystallization conditions or crystal formation that A-B-C subsites are preferred, since NMR [84] and enzymatic studies [85] suggested the occupancy of D subsite with (NAG)₃. HEWL Tyr62→His mutant showed enhanced catalysis of (NAG)₃ [86], which implied that (NAG)₃ might be binding deeper inside the pocket to enable breakdown of (NAG)₃ between subsites D and E. Powder diffraction [87] and MD simulation studies [83] also suggested binding of (NAG)₃ in B-C-D mode to be energetically more favorable. Powder diffraction showed occupancy of A-B-C and B-C-D subsites existing in a ratio of 35:65, implying that B-C-D mode might be favorable. Our calculations on buffalo LYZL proteins also suggested B-C-D subsites to be more favorable for binding of (NAG)₃. Analysis of hydrogen bonding network indicated that in B-C-D mode two or three more residues were involved in hydrogen bonding than in A-B-C binding mode in LYZL-(NAG)₃ (Table 2). Hydrogen bonding is an important contributor to protein stabilization [88] as well as protein-carbohydrate stability [43] and excess of hydrogen bonding could cooperatively stabilize LYZL-(NAG)₃ complex in the B-C-D binding mode.

Trp108 in HEWL and corresponding Trp in LYZL proteins are also completely conserved and known to make contact with substrate molecule (Table 1) as well as stabilize Glu35 through van der Waals interactions. Trp108 also contribute significantly to the overall stability of the molecule and Trp108→Tyr and Trp108→Gln mutants are reported to decrease not only conformational stability of the molecules but also catalytic activity and binding energy with the substrate molecules [76]. In human and HEWL, mutation of Arg114, also one of the residues involved in substrate binding, to His, Gln, Glu and Ala reduced the binding affinity [89–92]. In case of buffalo LYZL proteins, Arg114 is mutated to His in SLLP1, LYZL5 and LYZL6 and to Asn in LYZL2 and LYZL4. (NAG)₃ is a small molecule and it does not make contact with residue corresponding to position Arg114 and hence might not be of much consequence in the calculated binding energy with LYZL proteins. Similarly, residue number 37 (Asn 37 in HEWL) also did not show interaction with (NAG)₃ in both A-B-C and B-C-D binding modes and should not contribute in binding affinity toward LYZL proteins. This is consistent with the analysis of mutants Asn37→Gly and Asn37→Ser showing no significant change in binding constant with (NAG)₃ [2,93]. Nevertheless, these residues might be involved in interaction with larger receptors like molecules on oocytes. It is interesting to note that despite the large sequence divergence of LYZL proteins among themselves and from non-testicular lysozymes, many contacting residues in the binding pocket are conserved (Table 2). This could be necessitated for the conservation of functional binding efficacy of the LYZL proteins in sperm-oocyte or other cellular interactions.

Crystal structure analysis of mSLLP1 suggested the presence of glycan binding groove similar to that in lysozyme [61]; however, only partially conserved in the central cavity. The surface charges present in the groove also differed between the two structures. *In silico* docking suggested unfavorable interaction of mSLLP1 with (NAG)₄; however, binding with NAG and (NAG)₂ was preserved. In case of buffalo LYZL proteins, structural models showed binding with (NAG)₃ in two binding modes almost similar to those observed for non-testicular lysozymes. The mSLLP1 and human SLLP1 did not show significant binding with chitin or different glycans in a glycoarray [61]. The binding affinity of mSLLP1 and lysozyme also differed for glycan moieties. A new oocyte-specific receptor SAS1B has also been characterized that binds with sperm SLLP1 and helps in sperm-oocyte interaction [94]. These results indicated that LYZL proteins might have evolved for interaction with other receptors as well.

Evolution of LYZL proteins and their functional significance

Sequence analysis showed that LYZL proteins have undergone large divergence from each other and non-testicular lysozymes. Testicular *Lyzl* genes might have evolved from the ancestral genes through several gene duplications and accumulation of mutations to perform specific functions in the male reproductive system. The presence of most forms of LYZL proteins with high identity with orthologs but moderate with paralogs (38–50%) from diverse species suggested several events of duplication before the divergence of these species (Fig 3). Divergence might have resulted into loss or gain of function or both together as has been the case of evolution of α -lactalbumin from lysozymes [13,95]. Non-testicular lysozyme provide the innate immunity against the pathogens in various tissues or as a digestive enzyme in the stomach of ruminants, leaf-eating monkey and birds [5,7,8]. Testicular lysozyme family members might have evolved to facilitate reproduction [16,30,31] as well as protection against bacterial pathogens due to the presence of catalytic activity, e.g., LYZL6 has been observed biologically active [19].

The evolutionary pressure responsible for the existence of multiple forms of LYZL proteins and their large scale diversity in testis is not clear. Preincubation of mouse sperm with antibodies against SLLP1 [31] and LYZL4 proteins [30] reduced fertilization by 61% and 80%, respectively. The SLLP1 receptor SAS1B gene knockout female mice showed a decrease in fertilization only by 34% [94], which suggested that other factors are also important. It is also not clear if each member of LYZL proteins has evolved to carry out a specific function that might be supplemented by other members or not. The presence of similar set of LYZL proteins with high sequence similarity of orthologs among distant mammals suggests that each member might be playing important role in the reproduction or the protection of organisms. The localization of several of LYZL proteins on the surface of sperm suggest that these proteins might be involved in events mediated through surface contact. SLLP1 has been reported to interact with zona pellucida 3 protein [96] and implicated in sperm-oocyte binding [16,31].

Similarly, LYZL4 has also been implicated in fertility [30] and LYZL5 proposed as a biomarker of fertility [28]. Additionally, 10 extra c-terminal residues observed in LYZL5 shared 90% similarity with ankyrin repeat domain-containing protein 27. Ankyrin are known to be involved in many cellular events involving vesicle trafficking, vesicle fusion, molecular recognition and various signaling pathways [63]. The role of extra residues in LYZL5 is not discernible given a much smaller size of only 10 residues in comparison to 33 residues in an ankyrin repeat domain [62,63]. Nevertheless, the presence of extra c-terminal sequence showing high similarity with mammalian Varp protein in almost all LYZL5 proteins of mammalian species could be suggestive of its important role in sperm function.

The likelihood of the involvement of LYZL proteins in protection against pathogens seems weak, since had it been the case LYZL proteins would have been more efficient in providing

protection by being freely soluble in the seminal fluids rather than immobilized on the sperm surface. Moreover, functionally active lysozyme has also been detected in the seminal fluid to inhibit pathogens in testis [97] and thereby rendering LYZL proteins redundant in tissue protection. Nevertheless, the role of catalytically active LYZL proteins in innate immunity along with other function cannot be denied.

In SLLP1 of alligator and turtle, the catalytic residues are Glu35 and Asp52, which are similar to that present in catalytically active lysozymes. The SLLP1 of reptiles like python, cobra, garter snake and green anole are intermediate with only Asp52 replaced by Ser52, while Glu35 remaining conserved. Nevertheless, the SLLP1 in these species could be inactive as suggested by mutational analysis in active lysozymes [49]. The tissue-specific expression of SLLP1 in reptiles, and that if SLLP1 is a sperm-specific protein in reptiles is not known. The presence of only one *Lyzl* gene variant in reptiles or amphibians could be due to incomplete genome annotation or subsequent loss of some variants as has also been observed in some mammals [15]. It seems that in mammals SLLP1 and few other LYZL proteins diverged further and lost both the catalytic residues for specific function in reproduction. Sequence analysis suggests that SLLP1 in most of mammals is likely to be inactive because of mutations of both the residues. This has been shown in human [16] and buffalo (unpublished results). Unlike in mammals, there seems no study to implicate SLLP1 in the reproduction of reptiles. However, given the conserved catalytic residues in reptiles, it is not clear if the lack of catalytic activity in SLLP1 is not a pre-requisite for a specific role in reproduction or reptile SLLP1 underwent other adaptations to circumvent the requirement. The presence of conserved catalytic residues in reptiles also suggested that the loss of activity in SLLP1 should be a post speciation event. Similarly, the presence of some of the *Lyzl* genes in highly diverged extant vertebrates suggested that these paralogs should have diverged well before speciation of mammals from other vertebrates. However, it is not clear why LYZL2, which existed before the appearance of SLLP1 and LYZL5, have not been found in the genome of reptiles or frogs.

LYZL2 in almost all the mammalian species has conserved Glu35 and Asp52, except in platypus which has Ala52. These amino acid residues are encoded by GAA (Glu) and GAU (Asp) codons in milk and stomach lysozyme genes in buffalo and cattle. However, in buffalo LYZL2 and LYZL6, the corresponding codons are GAG and GAC identical to that observed in HEWL. The codon usage for the catalytic residues was also found conserved in human, mouse, cattle, and sheep LYZL2. In case of LYZL6, GAA (Glu35) is present in case of human and mouse instead of GAG as observed in buffalo, cattle and sheep, while GAC (Asp52) is conserved in all these species. Glu35 conserved in LYZL4 and LYZL5 is also encoded by GAG. It is noteworthy that all the conserved codons encoding the catalytically active residues (Glu35/Asp52) in buffalo LYZL proteins are identical to those observed in HEWL and not to the milk and stomach paralogs of buffalo. It suggests that the milk and stomach lysozymes have undergone evolutionary adaptation after duplication, while sperm lysozyme-like protein genes maintained the ancient codons identical to those present in distant vertebrates. In other cases, where the catalytic residues have been replaced, it was observed that in most of cases the change in amino acids was the result of a single or at the maximum two mutational events in the codons. Sequence variation also takes place to compensate a previous mutation to preserve the function [98]; nevertheless, large sequence variation is likely to cause some structural changes to enable specific function in the particular organ. In case of testis, the lysozyme type scaffold could have been selected by nature for molecular recognition of surface markers like glycan moieties on oocytes [31,96].

Supporting Information

S1 Table. Species wise list of lysozyme-like proteins taken in the present study.
(PDF)

Acknowledgments

SK was supported by the Junior and Senior Research Fellowships of Council for Scientific and Industrial Research, New Delhi, for doctoral program. The authors thank the Department of Biotechnology, Ministry of Science and Technology for providing computational facility under the framework of BTISnet.

Author Contributions

Conceptualization: JKK.

Data curation: SK JKK.

Formal analysis: SK JKK.

Funding acquisition: JKK.

Investigation: SK MAP AKM JKK.

Methodology: SK JKK.

Project administration: JKK.

Resources: JKK AKM.

Visualization: SK JKK.

Writing – original draft: SK MAP AKM JKK.

References

1. Jolles P, Jolles J (1984) What's new in lysozyme research? Always a model system, today as yesterday. *Mol Cell Biochem* 63: 165–189. PMID: [6387440](#)
2. Kumagai I, Sunada F, Takeda S, Miura K (1992) Redesign of the substrate binding site of hen egg white lysozyme based on the molecular evolution of C-type lysozymes. *J Biol Chem* 267: 4608–4612. PMID: [1537843](#)
3. McKenzie HA, White FH (1991) Lysozyme and α -lactalbumin: structure, function, and interrelationships. *Adv Protein Chem* 41: 173–315. PMID: [2069076](#)
4. Klockars M, Reitamo S (1975) Tissue distribution of lysozyme in man. *J Histochem Cytochem* 23: 932–940. PMID: [1104708](#)
5. Dobson DE, Prager EM, Wilson AC (1984) Stomach lysozyme of ruminants. *J Biol Chem* 259: 11607–11616. PMID: [6432801](#)
6. Irwin DM (1996) Molecular evolution of ruminant lysozymes. In: Jolles P (ed.), *Lysozymes: Model Enzymes in Biochemistry and Biology*. Basel: Birkhauser Verlag pp 347–364.
7. Stewart CB, Schilling JW, Wilson AC (1987) Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* 330: 401–404. doi: [10.1038/330401a0](#) PMID: [3120013](#)
8. Kornegay JR, Schilling JW, Wilson AC (1994) Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. *Mol Biol Evol* 11: 921–928. PMID: [7815930](#)
9. Irwin DM, Wilson AC (1989) Multiple cDNA sequences and the evolution of bovine stomach lysozyme. *J Biol Chem* 264: 11387–11393. PMID: [2738070](#)
10. Irwin DM, Wilson AC (1990) Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. *J Biol Chem* 265: 4944–4952. PMID: [2318875](#)
11. Yu M, Irwin D (1996) Evolution of stomach lysozyme: the pig lysozyme gene. *Mol Phylogenet Evol* 5: 298–308. doi: [10.1006/mpev.1996.0025](#) PMID: [8728388](#)
12. Phillips DC (1966) The three-dimensional structure of an enzyme molecule. *Sci Am* 215: 78–90.
13. McKenzie HA (1996) α -lactalbumins and lysozymes. In: Jolles P (ed.), *Lysozymes: Model Enzymes in Biochemistry and Biology*. Basel: Birkhauser Verlag, pp. 365–409.
14. Nitta K, Tsuge H, Shimazaki K, Sugai S (1988) Calcium-binding lysozymes. *Biol Chem Hoppe Seyler* 369: 671–675. PMID: [3214551](#)

15. Irwin DM, Biegel JM, Stewart CB (2011) Evolution of the mammalian lysozyme gene family. *BMC Evol Biol* 11: 1–16.
16. Mandal A, Klotz KL, Shetty J, Jayes FL, Wolkowicz MJ, Bolling LC, et al. (2003) SLLP1, a unique, intra-acrosomal, non-bacteriolytic, c lysozyme-like protein of human spermatozoa. *Biol Reprod* 68: 1525–1537. doi: [10.1095/biolreprod.102.010108](https://doi.org/10.1095/biolreprod.102.010108) PMID: [12606493](https://pubmed.ncbi.nlm.nih.gov/12606493/)
17. Zhang K, Gao R, Zhang H, Cai X, Shen C, Wu C, et al. (2005) Molecular cloning and characterization of three novel lysozyme-like genes, predominantly expressed in the male reproductive system of humans, belonging to the C-type lysozyme/alpha-lactalbumin family. *Biol Reprod* 73: 1064–1071. doi: [10.1095/biolreprod.105.041889](https://doi.org/10.1095/biolreprod.105.041889) PMID: [16014814](https://pubmed.ncbi.nlm.nih.gov/16014814/)
18. Narmadha G, Muneswararao K, Rajesh A, Yenugu S (2011) Characterization of a novel lysozyme-like 4 gene in the rat. *PLoS ONE* 6: e27659. doi: [10.1371/journal.pone.0027659](https://doi.org/10.1371/journal.pone.0027659) PMID: [22110709](https://pubmed.ncbi.nlm.nih.gov/22110709/)
19. Wei J, Li SJ, Shi H, Wang HY, Rong CT, Zhu P, et al. (2013) Characterisation of *Lyzls* in mice and anti-bacterial properties of human LYZL6. *Asian J Androl* 15: 1–7.
20. Guyonnet B, Moghaddam MZ, SanFrancisco S, Cornwall GA (2012) Isolation and proteomics characterization of the mouse sperm acrosomal matrix. *Mol Cell Proteomics* 11: 758–774. doi: [10.1074/mcp.M112.020339](https://doi.org/10.1074/mcp.M112.020339) PMID: [22707618](https://pubmed.ncbi.nlm.nih.gov/22707618/)
21. Wang G, Guo Y, Zhou T, Shi X, Yu J, Yang Y, et al. (2013) In-depth proteomic analysis of the human sperm reveals complex protein compositions. *J Proteomics* 79: 114–122. doi: [10.1016/j.jprot.2012.12.008](https://doi.org/10.1016/j.jprot.2012.12.008) PMID: [23268119](https://pubmed.ncbi.nlm.nih.gov/23268119/)
22. Gu B, Zhang J, Wu Y, Zhang X, Tan Z, Lin Y, et al. (2011) Proteomic analyses reveal common promiscuous patterns of cell surface proteins on human embryonic stem cells and sperms. *PLoS ONE* 6: e19386 doi: [10.1371/journal.pone.0019386](https://doi.org/10.1371/journal.pone.0019386) PMID: [21559292](https://pubmed.ncbi.nlm.nih.gov/21559292/)
23. Paasch U, Heidenreich F, Pursche T, Kuhlisch E, Kettner K, Grunewald S, et al. (2011) Identification of increased amounts of eppin protein complex components in sperm cells of diabetic and obese individuals by difference gel electrophoresis. *Mol Cell Proteomics* 10:1–14.
24. Amaral A, Castillo J, Estanyol JM, Balleca JL, Ramalho-Santos J, Oliva R (2013) Human sperm tail proteome suggests new endogenous metabolic pathways. *Mol Cell Proteomics* 12: 330–342. doi: [10.1074/mcp.M112.020552](https://doi.org/10.1074/mcp.M112.020552) PMID: [23161514](https://pubmed.ncbi.nlm.nih.gov/23161514/)
25. Chiu WWC, Erikson EK, Sole CA, Shelling AN, Chamley LW (2004). SPRASA, a novel sperm protein involved in immune-mediated infertility. *Hum Reprod* 19: 243–249. PMID: [14747161](https://pubmed.ncbi.nlm.nih.gov/14747161/)
26. Kovac JR, Pastuszak AW, Lamb DJ (2013) The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. *Fertil Steril* 99: 998–1007. doi: [10.1016/j.fertnstert.2013.01.111](https://doi.org/10.1016/j.fertnstert.2013.01.111) PMID: [23415969](https://pubmed.ncbi.nlm.nih.gov/23415969/)
27. McReynolds S, Dzieciatkowska M, Stevens J, Hansen KC, Schoolcraft WB, Katz-Jaffe MG (2014). Toward the identification of a subset of unexplained infertility: a sperm proteomic approach. *Fertil Steril* 102: 692–699. doi: [10.1016/j.fertnstert.2014.05.021](https://doi.org/10.1016/j.fertnstert.2014.05.021) PMID: [24934493](https://pubmed.ncbi.nlm.nih.gov/24934493/)
28. Agarwal A, Sharma R, Durairajanayagam D, Ayaz A, Cui Z, Willard B, et al. (2015) Major protein alterations in spermatozoa from infertile men with unilateral varicocele. *Reprod Biol Endocrinol* 13: 1–22.
29. Kwon WS, Rahman MS, Lee JS, Yoon S J, Park YJ, Pang MG (2015) Discovery of predictive biomarkers for litter size in boar spermatozoa. *Mol Cell Proteomics* 14: 1230–1240. doi: [10.1074/mcp.M114.045369](https://doi.org/10.1074/mcp.M114.045369) PMID: [25693803](https://pubmed.ncbi.nlm.nih.gov/25693803/)
30. Sun R, Shen R, Li J, Xu G, Chi J, Li L, et al. (2011) *Lyzl4*, a novel mouse sperm-related protein, is involved in fertilization. *Acta Biochim Biophys Sin* 43: 346–353. doi: [10.1093/abbs/gmr017](https://doi.org/10.1093/abbs/gmr017) PMID: [21444326](https://pubmed.ncbi.nlm.nih.gov/21444326/)
31. Herrero MB, Mandal A, Digilio LC, Coonrod SA, Maier B, Herr JC (2005) Mouse SLLP1, a sperm lysozyme-like protein involved in sperm-egg binding and fertilization. *Dev Biol* 284: 126–142 doi: [10.1016/j.ydbio.2005.05.008](https://doi.org/10.1016/j.ydbio.2005.05.008) PMID: [15982649](https://pubmed.ncbi.nlm.nih.gov/15982649/)
32. Wagner A, Holland OJ, Tong M, Shelling AN, Chamley LW (2014) The role of SPRASA in female fertility. *Reprod Sci* 10: 1–10.
33. Amaral A, Paiva C, Parrinello CA, Estanyol JM, Balleca JL, Ramalho-Santos J, et al. (2014) Identification of proteins involved in human sperm motility using high-throughput differential proteomics. *J Proteome Res* 13: 5670–5684. doi: [10.1021/pr500652y](https://doi.org/10.1021/pr500652y) PMID: [25250979](https://pubmed.ncbi.nlm.nih.gov/25250979/)
34. Puri P, Myers K, Kline D, Vijayaraghavan S (2008) Proteomic analysis of bovine sperm YWHA binding partners identify proteins involved in signaling and metabolism. *Biol Reprod* 79: 1183–1191. doi: [10.1095/biolreprod.108.068734](https://doi.org/10.1095/biolreprod.108.068734) PMID: [18753613](https://pubmed.ncbi.nlm.nih.gov/18753613/)
35. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2007) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1: 2856–2860.
36. Petersen TN, Brunak S, Von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786. doi: [10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701) PMID: [21959131](https://pubmed.ncbi.nlm.nih.gov/21959131/)

37. Sigrist CJA, de Castro E, Cerutti L, Cucho BA, Hulo N, Bridge A, et al. (2012) New and continuing developments at PROSITE. *Nuc Acids Res* 41: 1–4.
38. Edgar, Robert C (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acids Res* 32: 1792–1797.
39. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
40. Le SQ, Gascuel O (1993) An improved general amino acid replacement matrix. *Mol Biol Evol* 18: 691–699.
41. Krieger E, Vriend G (2015) New ways to boost molecular dynamics simulations. *J Comput Chem* 36: 996–1007. doi: [10.1002/jcc.23899](https://doi.org/10.1002/jcc.23899) PMID: [25824339](https://pubmed.ncbi.nlm.nih.gov/25824339/)
42. Gadave KS, Panda S, Singh S, Kalra S, Malakar D, Mohanty AK, et al. (2014) Structural and functional insights into the catalytic inactivity of the major fraction of buffalo milk xanthine oxidoreductase. *PLoS ONE* 9: e87618. doi: [10.1371/journal.pone.0087618](https://doi.org/10.1371/journal.pone.0087618) PMID: [24498153](https://pubmed.ncbi.nlm.nih.gov/24498153/)
43. Maenaka K, Matsushima M, Song H, Sunada F, Watanabe K, Kumagai I (1995) Dissection of protein-carbohydrate interactions in mutant hen egg white lysozyme complexes and their hydrolytic activity. *J Mol Biol* 247: 281–293. doi: [10.1006/jmbi.1994.0139](https://doi.org/10.1006/jmbi.1994.0139) PMID: [7707375](https://pubmed.ncbi.nlm.nih.gov/7707375/)
44. Cheatham JC, Artymuik PJ, Phillips DC (1992) Refinement of an enzyme complex with inhibitor bound at partial occupancy. Hen egg white lysozyme and tri-N-acetylchitotriose at 1.75 Å resolution. *J Mol Biol* 224: 613–628. PMID: [1569548](https://pubmed.ncbi.nlm.nih.gov/1569548/)
45. Ose T, Kuroki K, Matsushima M, Maenaka K, Kumagai I (2009) Importance of hydrogen bonding network including Asp52 for catalysis, as revealed by Asn59 mutant hen egg white lysozymes. *J Biochem* 146: 651–657. doi: [10.1093/jb/mvp110](https://doi.org/10.1093/jb/mvp110) PMID: [19605465](https://pubmed.ncbi.nlm.nih.gov/19605465/)
46. Harata K, Muraki M (1997) X-ray structure of turkey egg lysozyme complex with tri-N-acetyl-chitotriose. Lack of binding ability at subsite A. *Acta Crystallogr D Biol Crystallogr* 53: 650–657. doi: [10.1107/S0907444997005362](https://doi.org/10.1107/S0907444997005362) PMID: [15299852](https://pubmed.ncbi.nlm.nih.gov/15299852/)
47. Karlsen S, Hough E (1995) Crystal structures of three complexes between chito-oligosaccharides and lysozyme from the rainbow trout. How distorted is the NAG sugar in site D? *Acta Crystallogr D Biol Crystallogr* 51: 962–978. doi: [10.1107/S0907444995005105](https://doi.org/10.1107/S0907444995005105) PMID: [15299765](https://pubmed.ncbi.nlm.nih.gov/15299765/)
48. Vollan VB, Hough E, Karlsen S (1999) Structural studies on the binding of 4-methylumbelliferone glycosides of chitin to rainbow trout lysozyme. *Acta Crystallogr D Biol Crystallogr* 55: 60–66. doi: [10.1107/S0907444998006623](https://doi.org/10.1107/S0907444998006623) PMID: [10089395](https://pubmed.ncbi.nlm.nih.gov/10089395/)
49. Hadfield AT, Harvey DJ, Archer DB, MacKenzie DA, Jeenes DJ, Radford SE, et al. (1994) Crystal structure of mutant D52S hen egg white lysozyme with an oligosaccharide product. *J Mol Biol* 243: 856–872. doi: [10.1006/jmbi.1994.1688](https://doi.org/10.1006/jmbi.1994.1688) PMID: [7966306](https://pubmed.ncbi.nlm.nih.gov/7966306/)
50. Kawaguchi Y, Yoneda K, Torikata T, Araki T (2015) Asp48 function in the hydrogen-bonding network involving Asp52 of hen egg-white lysozyme. *Biosci Biotechnol Biochem* 79: 196–204. doi: [10.1080/09168451.2014.963502](https://doi.org/10.1080/09168451.2014.963502) PMID: [25514638](https://pubmed.ncbi.nlm.nih.gov/25514638/)
51. Song H, Inaka K, Maenaka K, Matsushima M (1994) Structural changes of active site cleft and different saccharide binding modes in human lysozyme co-crystallized with hexa-N-acetyl-chitohexaose at pH 4.0. *J Mol Biol* 244: 522–540. doi: [10.1006/jmbi.1994.1750](https://doi.org/10.1006/jmbi.1994.1750) PMID: [7990138](https://pubmed.ncbi.nlm.nih.gov/7990138/)
52. Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc Acids Res* 25: 3389–3402.
53. Krieger E, Joo K, Lee J, Raman S, Thompson J (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 77: 114–122. doi: [10.1002/prot.22570](https://doi.org/10.1002/prot.22570) PMID: [19768677](https://pubmed.ncbi.nlm.nih.gov/19768677/)
54. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins* 64: 559–574. doi: [10.1002/prot.20921](https://doi.org/10.1002/prot.20921) PMID: [16736488](https://pubmed.ncbi.nlm.nih.gov/16736488/)
55. Krieger E, Nielsen JE, Spronk CA, Vriend G (2006) Fast empirical pKa prediction by Ewald summation. *J Mol Graph Model* 25: 481–486. doi: [10.1016/j.jmgm.2006.02.009](https://doi.org/10.1016/j.jmgm.2006.02.009) PMID: [16644253](https://pubmed.ncbi.nlm.nih.gov/16644253/)
56. Krieger E, Dunbrack RL Jr, Hooft RW, Krieger B (2012) Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods Mol Biol* 819: 405–421. doi: [10.1007/978-1-61779-465-0_25](https://doi.org/10.1007/978-1-61779-465-0_25) PMID: [22183550](https://pubmed.ncbi.nlm.nih.gov/22183550/)
57. Trott O, Olson AJ (2010) AutoDock: improving the speed and accuracy of docking with a new scoring function, efficient, optimization and multithreading. *J Comput Chem* 31: 455–461. doi: [10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334) PMID: [19499576](https://pubmed.ncbi.nlm.nih.gov/19499576/)
58. Blake CCF, Johnson LN, Mair GA, North ACT, Phillips DC, Sarma VR (1967) Crystallographic studies of the activity of hen egg white lysozyme. *Proc R Lond Ser B Biol Sci* 167: 378–388.

59. Imoto T, Johnson LN, North ACT, Phillips DC, Rupley JA (1972) Vertebrate lysozymes. In Boyer P.D. (ed.), *The Enzymes*, Academic Press, New York, 7: pp. 665–868.
60. Kumagai I, Maenaka K, Sunada F, Takeda S, Miura K (1993) Effects of subsite alterations on substrate-binding mode in the active site of hen egg-white lysozyme. *Eur J Biochem* 212: 151–156. PMID: [8444153](#)
61. Zheng H, Mandal A, Shumilin A, Chordia MD, Panneerdoss S, Herr JC, et al. (2015) Sperm Lysozyme-Like Protein 1 (SLLP1), an intra-acrosomal oolemmal-binding sperm protein, reveals filamentous organization in protein crystal form. *Andrology* 3: 756–771. doi: [10.1111/andr.12057](#) PMID: [26198801](#)
62. Schafer IB, Hesketh GG, Bright NA, Gray SR, Pryor PR, Evans PR, Luzio JP, Owen DJ (2012) The binding of Varp to VAMP7 traps VAMP7 in a closed, fusogenically inactive conformation. *Nat Struct Mol Biol* 19: 1300–1309. doi: [10.1038/nsmb.2414](#) PMID: [23104059](#)
63. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The Ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* 13: 1435–1448. doi: [10.1110/ps.03554604](#) PMID: [15152081](#)
64. Maenaka K, Matsushima M, Kawai G, Kidera A, Watanabe K, Kuroki R, et al. (1998) Structural and functional effect of Trp62→Gly and Asp101→Gly substitutions on substrate-binding modes of mutant hen egg-white lysozymes. *Biochem J* 333: 71–76. PMID: [9639564](#)
65. Garcia-Hernandez E, Zubillaga RA, Chavelas-Adame EA, Vazquez-Contreras E, Rojo-Dominguez A, Costas M (2003) Structural energetics of protein-carbohydrate interactions: Insights derived from the study of lysozyme binding to its natural saccharide inhibitors. *Protein Sci* 12: 135–142. doi: [10.1110/ps.0222503](#) PMID: [12493836](#)
66. Martinez-Heredia J, Estanyol JM, Balleca JL, Oliva R. (2006) Proteomic identification of human sperm proteins. *Proteomics* 6: 4356–4369. doi: [10.1002/pmic.200600094](#) PMID: [16819732](#)
67. de Mateo S, Martinez-Heredia J, Estanyol JM, Dominguez-Fandos D, Vidal-Taboada JM, Balleca J L, et al. (2007) Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. *Proteomics* 7: 4264–4277. doi: [10.1002/pmic.200700521](#) PMID: [18040982](#)
68. Baker MA, Reeves G, Hetherington L, Muller J, Baur I, Aitken RJ (2007) Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics Clin Appl* 1: 524–532. doi: [10.1002/prca.200601013](#) PMID: [21136703](#)
69. Peddinti D, Nanduri B, Kaya A, Feugang JM, Burgess SC, Memili E (2008) Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Syst Biol* 2: 1–13.
70. Saadi HAS, Riemsdijk EV, Dance AL, Rajamanickam GD, Kastelic JP, Thundathil JC (2013) Proteins associated with critical sperm functions and sperm head shape are differentially expressed in morphologically abnormal bovine sperm induced by scrotal insulation. *J Proteomics* 82: 64–80. doi: [10.1016/j.jprot.2013.02.027](#) PMID: [23500133](#)
71. Swegen A, Curry BJ, Gibb Z, Lambourne SR, Smith ND, Aitken RJ (2015) Investigation of the stallion sperm proteome by mass spectrometry. *Reproduction* 3: 235–244.
72. Baker MA, Naumovski N, Hetherington L, Weinberg A, Velkov T, Aitken RJ (2013) Head and flagella subcompartmental proteomic analysis of human spermatozoa. *Proteomics* 13: 61–74. doi: [10.1002/pmic.201200350](#) PMID: [23161668](#)
73. de Mateo S, Castillo J, Estanyol JM, Balleca JL, Oliva R (2011) Proteomic characterization of the human sperm nucleus. *Proteomics* 11: 2714–2726. doi: [10.1002/pmic.201000799](#) PMID: [21630459](#)
74. Byrne K, Leahy T, McCulloch R, Colgrave ML, Holland MK (2012) Comprehensive mapping of the bull sperm surface proteome. *Proteomics* 12: 3559–3579. doi: [10.1002/pmic.201200133](#) PMID: [23081703](#)
75. Muraki M, Harata K, Hayashi Y, Machida M, Jigami Y (1991) The importance of precise positioning of negatively charged carboxylate in catalytic action of human lysozyme. *Biochim Biophys Acta* 1079: 229–237. PMID: [1911846](#)
76. Inoue M, Yamada H, Yasukochi T, Kuroki R, Miki T, Horiuchi T, et al. (1992). Multiple role of hydrophobicity of tryptophan-108 in chicken lysozyme: structural stability, saccharide binding ability, and abnormal pKa of glutamic acid-35. *Biochemistry* 31: 5545–5553. PMID: [1610799](#)
77. Mulvey RS, Gualtieri RJ, Beychok S (1973) Spectral properties of human lysozyme and its inhibitor complexes. Fluorescence and difference spectra. *Biochemistry* 12: 2683–2690. PMID: [4514395](#)
78. Mulvey RS, Gualtieri RJ, Beychok S (1974) Composition, fluorescence and circular dichroism of rat lysozyme. *Biochemistry* 13: 782–787. PMID: [4521116](#)
79. Kumagai I, Miura K (1989) Enhanced bacteriolytic activity of hen egg white lysozyme due to conversion of Trp62 to other aromatic amino acid residues. *J Biochem* 105: 946–948. PMID: [2549023](#)
80. Kumagai I, Kojima S, Tamaki E, Miura K (1987) Conversion of Trp 62 of hen egg white lysozyme to Tyr by site directed mutagenesis. *J Biochem* 102: 733–740. PMID: [3325501](#)

81. Maenaka K, Kawai G, Watanabe K, Sunada F, Kumagai I (1994) Functional and structural role of a tryptophan generally observed in protein-carbohydrate interaction. TRP-62 of hen egg white lysozyme. *J Biol Chem* 269: 7070–7075. PMID: [8125914](#)
82. Muraki M, Harata K, Sugita N, Sato K (2000) Protein-carbohydrate interactions in human lysozyme probed by combining site-directed mutagenesis and affinity labeling. *Biochemistry* 39: 292–299. PMID: [10630988](#)
83. Zhong Y, Patel S (2013) Binding structures of tri-N-acetyl- β -glucosamine in hen egg white lysozyme using molecular dynamics with a polarizable force field. *J Comput Chem* 34: 163–174. doi: [10.1002/jcc.23109](#) PMID: [23109228](#)
84. Lumb KJ, Cheatham JC, Dobson CM (1994) ^1H nuclear magnetic resonance studies of hen lysozyme-N-acetylglucosamine oligosaccharide complexes in solution. Application of chemical shifts for the comparison of conformational changes in solution and in the crystal. *J Mol Biol* 235: 1072–1087. doi: [10.1006/jmbi.1994.1058](#) PMID: [8289309](#)
85. Rupley JA (1967) The binding and cleavage of lysozyme of N-acetylglucosamine oligosaccharides. *Proc R Soc Lond B Biol Sci* 167: 416–428. PMID: [4382804](#)
86. Maenaka K, Matsushima M, Kawai G, Watanabe K, Kuroki R, Kumagai I (1998) Structural analysis of mutant hen egg-white lysozyme preferring a minor binding mode. *Biochim Biophys Acta* 84: 23–31
87. Von Dreele RB (2005) Binding of N-acetylglucosamine oligosaccharides to hen egg-white lysozyme: a powder diffraction study. *Acta Crystallogr D Biol Crystallogr* 61: 22–32. doi: [10.1107/S0907444904025715](#) PMID: [15608372](#)
88. Kaushik JK, Imura S, Ogasahara K, Yamagata Y, Segawa S, Yutani K (2006) Completely buried, non-ion-paired glutamic acid contributes favorably to the conformational stability of pyrrolidone carboxyl peptidases from hyperthermophiles. *Biochemistry* 45: 7100–7112. doi: [10.1021/bi052610n](#) PMID: [16752900](#)
89. Harata K, Muraki M, Jigami Y (1993) Role of Arg115 in the Catalytic Action of Human Lysozyme: X-ray Structure of His115 and Glu115 Mutants. *J Mol Biol* 233: 524–535. doi: [10.1006/jmbi.1993.1529](#) PMID: [8105095](#)
90. Muraki M, Morikawa M, Jigami Y, Tanaka H (1989) A structural requirement in the subsite F of lysozyme: The role of arginine 115 in human lysozyme revealed by site-directed mutagenesis. *Eur J Biochem* 179: 573–579. PMID: [2493374](#)
91. Toshima G, Kawamura S, Araki T, Torikata T (2003) Histidine-114 at subsites E and F can explain the characteristic enzymatic activity of guinea hen egg-white lysozyme. *Biosci Biotechnol Biochem* 67: 540–546. doi: [10.1271/bbb.67.540](#) PMID: [12723601](#)
92. Kawamura S, Chijiwa Y, Minematsu T, Fukamizo T, Varum KM, Torikata T (2008) The role of Arg114 at subsites E and F in reactions catalysed by hen egg white lysozyme. *Biosci Biotechnol Biochem* 72: 823–832. doi: [10.1271/bbb.70694](#) PMID: [18323666](#)
93. Kawamura S, Eto M, Imoto T, Ikemizu S, Araki T, Torikata T (2004) Functional and structural effects of mutagenic replacement of Asn37 at subsite F on the lysozyme-catalysed reaction. *Biosci Biotechnol Biochem* 68: 593–601. doi: [10.1271/bbb.68.593](#) PMID: [15056892](#)
94. Sachdev M, Mandal A, Mulders S, Digilio LC, Panneerdoss S, Suryavathi V, et al. (2012) Oocyte specific oolemmal SAS1B involved in sperm binding through intra-acrosomal SLLP1 during fertilization. *Dev Biol* 363: 40–51. doi: [10.1016/j.ydbio.2011.12.021](#) PMID: [22206759](#)
95. Qasba PK, Kumar S (1997) Molecular divergence of lysozymes and alpha-lactalbumin. *Crit Rev Biochem Mol Biol* 32: 255–306. doi: [10.3109/10409239709082574](#) PMID: [9307874](#)
96. Naz RK, Dhandapani L (2010) Identification of human sperm proteins that interact with human zona pellucida3 (ZP3) using yeast two-hybrid system. *J Reprod Immunol* 84: 24–31. doi: [10.1016/j.jri.2009.10.006](#) PMID: [19945174](#)
97. Mardh PA, Colleen S (1974) Lysozyme in seminal fluid of healthy males and patients with prostatitis and in tissues of male urogenital tract. *Scand J Urol Nephrol* 8: 179–183. PMID: [4610730](#)
98. Baldwin E, Xu J, Hajiseyedjavadi O, Baase WA, Matthews BW (1996) Thermodynamic and structural compensation in “size-switch” core repacking variants of bacteriophage T4 lysozyme. *J Mol Biol* 259: 542–559. doi: [10.1006/jmbi.1996.0338](#) PMID: [8676387](#)