Applied Biological Research 18 (2): 171-177; (2016)
DOI: 10.5958/0974-4517.2016.00026.4

INSILICO CHARACTERIZATION OF NOVEL PH20 HYALURONIDASE PROTEIN SEQUENCES OF MURRAH *(Bubalus bubalis)* BULLS

Poonam Sikka*, Abhigyan Nath and Keerti Kumar Yadav

Central Institute for Research on Buffaloes, Hisar – 125 001, Haryana (India)
*e-mail: *drsikapunam@gmail.com*

(Received 13 April, 2016; Accepted 18 June, 2016)

ABSTRACT

PH20 hyaluronidase proteins are members of Zona pellucida binding protein family. PH20 hyaluronidase is reportedly bifunctional i.e. have hyaluronidase and sperm-zona binding activities. In this study, we performed insilico characterization of novel PH20 protein using partial nucleotide sequences, obtained from SNP genotyping of *Bubalus bubalis* bulls, by Sanger's dideoxy sequencing method. These gene sequences are novel and first time reported in *B. bubalis*. Possible glycosylation and mannosylation patterns were also predicted in both the PH20 hyaluronidase sequences which appear responsible for its hyaluronidase activity. Further, the potential tertiary structure of PH20 was deduced using Swiss-Model server. The tertiary structure models of PH20 were validated by ProSA server and Ramachandran plots. The tertiary structure models of PH20 were further refined using KoBaMIN server and validated by PROCHECK, ProSA and Ramachandran plots. These structural models were assigned PMDB identifiers PM0080458 and PM0080459. Physico-chemical characterization revealed that *B. bubalis* PH20 hyaluronidase are more thermostable and more hydrophobic than that of cattle PH20 hyaluro-nidase. The functional annotations were further inferred using Gene Ontolgy based predictions. Present work on elucidated structure of PH20 predicts a reference model for functional genomics studies related to multifunctionality of PH20 hyaluronidase proteins with respect to genetic variants within the species or trans-species. The current work is the first preliminary insilico study on novel bubaline PH20 hyaluronidases.

Key words: Gene ontology, glycosylation, homology modeling, PH20 hyaluronidase, zona pellucida binding protein

INTRODUCTION

A wide variety of proteins designated as zona pellucida binding proteins (ZPBP) are involved in sperm-egg interaction (McLeskey et al.,1998). ZPBP group includes PH20 hyaluronidase and several CUB domain (Bork and Beckmann,1993) comprising proteins as spermadhesins. Spermadhesins constitute a bulk of seminal plasma proteins, all comprising of 110-133 amino acid residues, organized in β-stands – a character of CUB domain. PH20 has hyaluronidase activity which facilitates sperm penetration and fertilization and sperm-zona binding (Yudin *et al.*, 1999; Cherr *et al.*, 2001; Jansen *et al.*, 2001). PH20 belongs to the glycoside hydrolase 56 protein family (Henrissat *et al.*, 1995) which is a group of hyaluronidases that are involved in the hydrolysis of glycosidic bonds present in sperm membrane. It is located both on acrosomal membrane and sperm plasma membrane (Tung *et al.*, 1997). In present work, the tertiary structure of PH20 hyaluronidase is elucidated by homology modeling, based on novel (first report in buffaloes) sequences

(Accession No. KU364415 and KU364416) obtained from SNP genotyping of Murrah (*Bubalus bubalis*) bulls and subsequent Sanger's sequencing to identify genetic variance for fertility. Physicochemical properties, determined for predicted tertiary structures based on polymorphic novel PH20 sequences identified in buffalo bulls, has been attempted to ascribe multifunctional annotation to polymorphic novel DNA sequences of PH20 protein.

MATERIALS AND METHODS

Targeted sequences for structural modeling

KU364415 and KU364416 (GeneBank) are novel polymorphic nucleotide sequences of partial coding region of PH20, obtained through SNP genotyping of buffalo (*Bubalus bubalis*) bulls. Two nucleotide sequences were obtained by Sanger's dideoxy sequencing of partially amplified gene which generated 221 and 189 amino acid residues, respectively.

Insilico characterization

Presence of protein domains in the novel *Bubalus bubalis* PH20 hylauronidase sequences was inferred through follow-up search by CDD (Conserved domain database) (Marchler-Bauer *et al.*, 2005) and InterProScan (Quevillon *et al.*, 2005). Glycosylation predictor server (Hamby and Hirst, 2008) and NetCGlyc 1.0 (Julenius, 2007) were used to predict glycosylation sites and C-mannosylation sites, respectively. Sequence-based identification and further analysis was done using SIAS tool (*http://imed.med.ucm.es/Tools/sias.html*). The basic physicochemical properties of the sequences were calculated by ProtParam tool (Gasteiger *et al.*, 2005). Gene ontology terms were predicted using protein function prediction server (Khan *et al.*, 2015).

Homology modeling and analysis

Homology modeling structure prediction for PH20 protein, deduced from novel identified nucleotide sequences [KU364415 and KU36416] was performed using Swiss-Model Suite of Expasy platform (Biasini *et al.*, 2014). The initial homology models thus obtained were further refined using KoBaMIN server (Rodrigues *et al.*, 2012); i) using knowledge based potential for energy minimization and ii) streochemistry correction. Validation of refined modelled theoretical structures was attempted using PROCHECK (Laskowski *et al.*, 1993). Ramachandran plots, created by using RAMPAGE tool (*http://mordred.bioc.cam.ac.uk/~rapper/rampage.php*) and Z-score plots obtained from ProSA server (Wiederstein and Sippl, 2007) were generated to establish authenticity for the modeled structures. The two modeled structures of PH20 protein were compared using TM-align server (Zhang and Skolnick, 2005).

RESULTS AND DISCUSSION

CDD search and InterProScan confirmed the presence of glycoside hydrolase activity and glycoside hydrolase 56 domain in the deduced protein models generated from Sanger's sequencing of bull genotyping products for PH20 gene. Glycoside hydrolase family 56 has been recognized as having hyaluronidase enzyme-like activity (*http://www.ebi.ac.uk/interpro/entry/IPR018155*) out of 85 different families of glycosyl hydrolases classified on basis of sequence similarities.

The percent similarity matrix was determined among protein sequences of ZPBP family members as XP_586790.5-*Bos taurus Predicted PH20 Hyaluronidase*, NP_001008413.3-*Bos taurus* Sperm Adhesin Molecule 1 Precursor, AAM60770.1-*Homo sapiens* Hyaluronidase1 variant 1,

Table 1: Percent similarity (%) of *Bubalus bubalis* PH20 nucleotide sequences with other protein members of ZPBP family

Nucleotide sequences	KU36445	KU364416	XP_586790.5	NP_001008413.3	AAM60770.1	XP_005205168.1	XP_002703276.3
KU36445	100.00						
KU364416	84.65	100.00					
XP_58790.5	82.35	80.42	100.00				
NP_001008413.3	68.32	63.49	69.09	100.00			
AAM60770.1	42.08	40.21	38.27	39.01	100.00		
XP_005205168.1	12.21	10.58	12.28	11.14	10.00	100.00	
XP_002703276.3	8.96	8.27	10.34	9.65	8.27	8.27	100.00

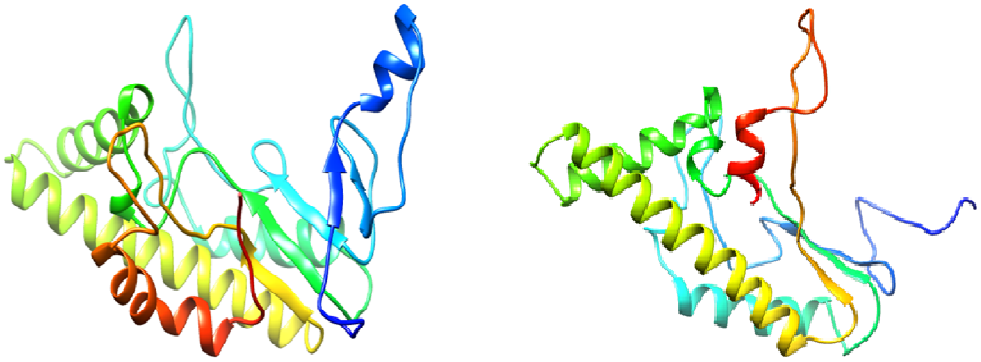

Fig. 1: Homology models of KU364415 (left) and KU364416 (b) of *Bubalus bubalis* PH20 as predicted by Swiss-Model server

XP_005205168.1-*Bos taurus* ZPBP isoform X1, XP_002703276.3-*Bos taurus* AQN1 and the PH20 hyaluronidase protein sequences deduced first time reported novel sequences from partial gene of PH20 hyaluronidase from *Bubalus bubalis* (KU36441 and KU364416) by using SIAS tool (Table 1). It indicates that the *Bubalus bubalis* PH20 hyaluronidase sequences are more close to the *Bos taurus* PH20 hyaluronidase sequences than to any other member of ZPBP family. The sequence diversity is very pronounced in ZPBP family proteins specifically with respect to PH20 and other ZPBP molecules as Spermadhesins (AQN-1) [Ekhlasi-Hundrieser *et al.*, 2002].

Homology modeling of *B. bubalis* PH20 hyaluronidase (deduced from KU364415 and KU364416) has been attempted using Swiss-Model server of Expasy platform (*http://swissmodel.expasy.org*) [Fig. 1a,b]. For both the *B. bubalis* PH20 sequences, the best template identified was of human hyaluronidase 1 (Protein DataBankID-2PE4) which had 42.27 and 40.11% similarity with the deduced protein sequences from KU364415 and KU364416, respectively. The initial Swiss-Model homology model structures were further refined using KoBaMIN server and further validated by PROCHECK. A g-factor value of more than -0.5, qualifies both the predicted modeled structures as structurally stable.

The presence of potential glycosylation sites in both the PH20 protein sequences were determined using Glycosylation prediction server which is based on random forest algorithm and pair-wise patterns (Hamby and Hirst, 2008). Post-translational modifications introduce functional diversity in proteins. The presence of potential glycosylation sites in PH20 have previously been reported (Lathrop, 1990). C-mannosylation modulates protein activity and stability as cited in case of human hyaluronidase 1 (Goto *et al.*, 2014). The potential glycosylation and C-mannosylation sites are predicted in buffalo PH20 (KU364415 and KU364416) as well (Table 2). TM-alignment of two predicted aforesaid models has been attempted to study the structural variation between the two theoretical models. TM-align is based on TM-score (template matching score) rotation and dynamic programming, which is used to assess the quality of modeled protein structures. A TM score between 0.5 and 1.0 represents structural similarity between the compared models. TM-scores of 0.80 and 0.99 were obtained while using KU364415 and KU364416 as reference sequence, respectively.

Table 2: Predicted glycosylated and C-mannosylated sites in PH20 protein sequences of *B. bubalis*

	Predicted glycosylated sites (position - glycosylated amino acid residue)
KU364415	3-T,4-N,32-S,61-N,63-S,134-N,161-T,188-N,193-N,214-S,220-S
KU364416	13-S,40-N,115-N,124-T,142-T,161-N,181-N
	Predicted C-mannosylated sites (position-C-mannosylated amino acid residue)
KU364415	105-W
KU364416	86-W

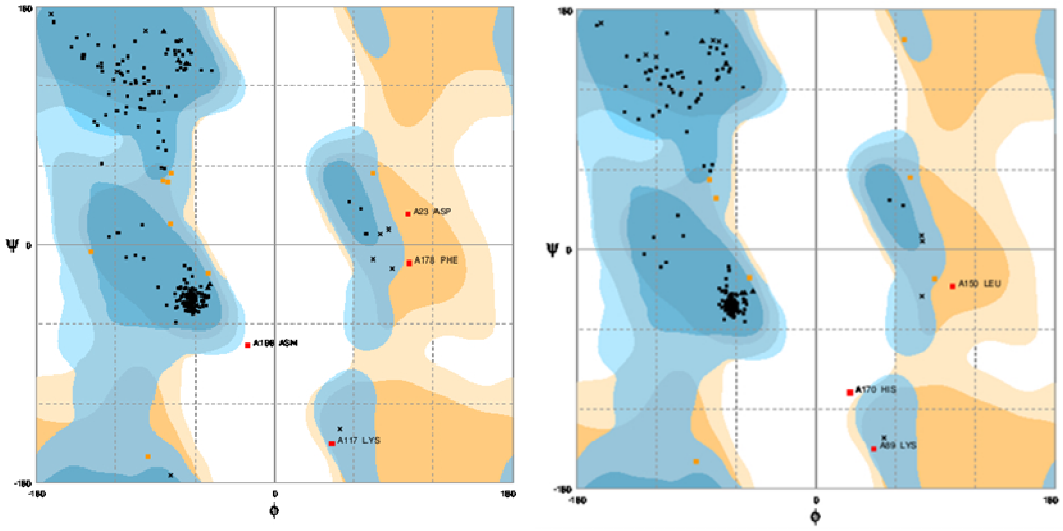

Fig. 2: Ramachandran plot for KU364415 (left side) and KU364416 (right side)

Ramachandran plots, prepared by using Rampage tool validate the theoretical models by observing various residues falling in the allowed and disallowed regions of plot, depicted that most of the residues fall under the allowed regions, indicating the stability of the modeled protein structures (Fig. 2). The quality of the predicted theoretical models was also checked/validated through ProSA webserver. The Z-score which defines the overall model quality was determined by comparing with Z-scores of the experimentally determined (X-ray, NMR) structures of similar size selected from the Protein Data Bank. ProSA-Z-score plots for the predicted protein model are given in Fig. 3. Study infers that the determined Z-scores for the predicted models are in acceptable range which in turn reveals that the predicted model structures are structurally stable. The modeled structures have been submitted to Protein Model Database (Castrignanò *et al.*, 2006) with identifiers: PM0080507 and PM0080508 for KU364415 and KU364416, respectively**.**

The various physicochemical properties of both PH20 hyaluronidase sequences are given in Table 3. The deduced protein sequences from KU364415 and KU364416 consisting of 221 and 189 amino acid residues respectively, have estimated molecular weight of 25868.5 and 22109.2 Da. Estimated Pi is comparable with other bovidae family members which is greater than the human hyaluronidase. The instability index (Guruprasad *et al.*, 1990) values of *Bubalus bubalis* PH20 hyaluronidase falls between instability index given for human and cattle, indicating that buffalo PH20 structures are more stable than the cattle PH20 hyaluronidase.

Aliphatic index of a globular protein is related to its thermostability and is determined by the volume occupied by aliphatic amino acids (leucine, isoleucine, valine and alanine) [IKAI, 1980]. Higher aliphatic index of *B. bubalis* as determined in this study indicated that they are more thermo-stable than that of cattle PH20 hyaluronidase. The GRAVY scores (Kyte and Doolittle, 1982) for *B. bubalis* PH20 sequences are higher than that of both *Homo-sapiens* and *Bos taurus* hyaluronidase proteins which indicates more hydrophobicity associated with bubaline protein structures. The difference in physicochemical properties between the two *B. bubalis* sequences is attributed to the

Table 3: The physicochemical properties of PH20 hyaluronidase of *Bubalus bubalis*, *Homo sapiens* and *Bos taurus*

Properties	KU364415	KU364416	*Homo sapiens* hyaluronidase1	*Bos taurus* hyaluronidase
Number of amino acids	221	189	424	553
Molecular weight(Da)	25868.5	22109.2	47367.6	62281.1
Theoretical Pi	8.61	8.92	6.98	8.79
Total number of negatively charged residues (Asp + Glu)	26 (11%)	23 (12%)	34 (8%)	46(8.31%)
Total number of positively charged residues (Arg + Lys)	29(13%)	28(14.8%)	33(7.7 %)	57 (10.3%)
Instability index	40.07	45.43	38.57	50.55
Aliphatic index	82.53	69.68	74.83	50.55
Grand average of hydropathicity (GRAVY)	-0.523	-0.788	-0.259	-0.303

polymorphic nature of the two sequences obtained from two individuals of the same species. Also, the partial sequences may possibly have variant amino acid residues in final protein resulting from SNPs in the nucleotide sequences that may result in phenotypic differences related to reproduction.

Gene Ontology (GO) provides unified controlled vocabularies that can be used to represent gene and gene products across all organisms. The GO terms for deduced protein sequences are given in Table 4 (only high scoring GO terms are shown). The GO terms are similar for both buffalo PH20 sequences. The GO terms 0003824, 0004415, 0015929, 0033906 and 0005975 indicate towards the hyaluronidase activity; while GO term 0005515 indicates towards the protein-protein interaction/binding of the sequences which may be due to the possible role of PH20 hyaluronidase in sperm-zona binding. The predicted gene ontology terms indicated towards both enzymatic activity as well as protein binding.

The presence of glycoside hydrolase 56 domain detected in PH20 using CDD search and InterProScan in our models indicates glycan moieties associated to PH20 sequences (Acc. No. KU364415 and KU364416), which may be responsible for its binding and hyaluronidase activity. Physico-chemical characterization revealed that the *B. bubalis* PH20 hyaluronidases are more stable and more hydrophobic than those of cattle PH20 hyaluronidase. The elucidated PH20 sequences i.e. KU364415 and KU364416 are polymorphic in *Bubalus bubalis* Murrah bulls. Using these sequences homology modeling of protein has been attempted in reference to human hyaluronidase (2PE4). The modeled structures were refined using KoBaMIN server and were tested for stereochemical

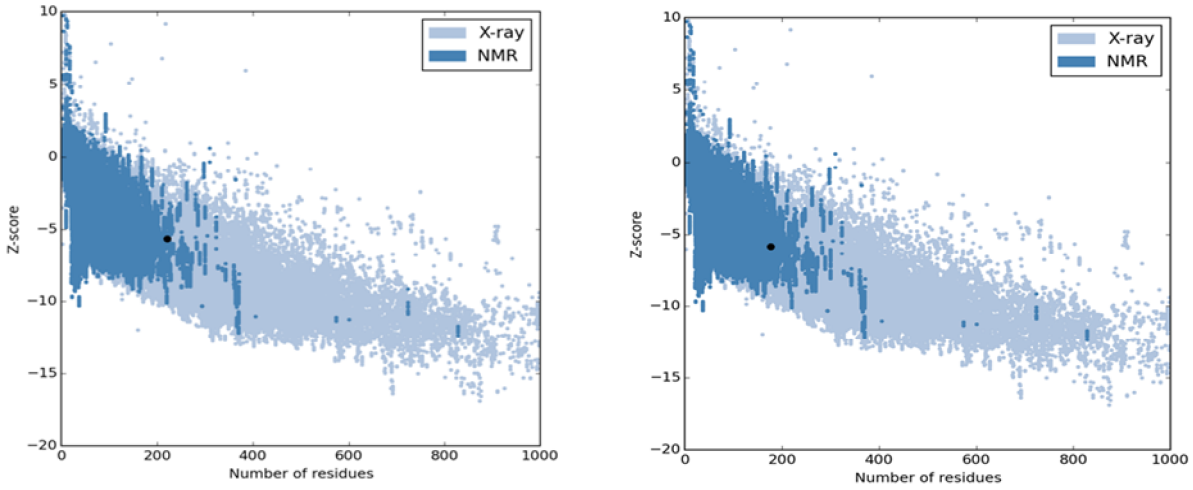

Fig. 3: ProSA–Z-score plot for KU364415 (left side) and (b) KU364416 (right side)

Table 4: Gene ontology predictions for PH20 hyaluronidase of *Bubalus bubalis*

GO components	Term-description
Molecular function terms	GO:0003824-catalytic activity
	GO:0004415-hyalurononglucosamindase activity
	GO:0015929-hexoaminidase activity
	GO:0005515-protein binding
	GO:0033906-hyalurononglucosamindase activity
Biological process terms	GO:0005975-carbohydrate metabolic process
	GO:0030214-hyaluronon catabolic process
Cellular component terms	GO:0005576-extracellular region

quality using PROCHECK, Ramachandran plots and ProSAserver. The predicted models qualified the criteria of stability. These structural models are registered in PMDB with identifiers: PM0080507 and PM0080508. It is the first report of structural model of nucleotide sequences based insilico designed PH20 proteins in Bubaline species. Although both the *Bubalus bubalis* PH20 sequences are polymorphic as indicated from the sequence comparison analysis, the overall structure and function are preserved.

Polymorphic characters of sequences are attributed to multi-functionality in response to ambient/environment changes either due to difference in reproductive efficiency or from one tissue to other within reproductive tract due to tissue based regulation of differential protein expression as a temporal effect. Polymorphism ascribed to regulatory proteins show up as one of the promising markers for male fertility index determination. Current deduced structural models are a resource for advance studies on bull fertility genomics.

REFERENCES

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. and Schwede, T. 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**(W1): W252-W258.

Bork, P. and Beckmann, G. 1993. The CUB Domain. *Journal of Molecular Biology*, **231**: 539-545.

Castrignanò, T., De Meo, P.D.O., Cozzetto, D., Talamo, I.G. and Tramontano, A. 2006. The PMDB protein model satabase. *Nucleic Acids Research,* **34**(suppl 1): D306-D309.

Cherr, G.N., Yudin, A.I. and Overstreet, J.W. 2001. The dual functions of GPI-anchored PH-20: Hyaluronidase and intracellular signaling. *Matrix Biology*, **20**: 515-525.

Ekhlasi-Hundrieser, M., Sinowatz, F., De Wilke, I.G., Waberski, D. and Töpfer-Petersen, E. 2002. Expression of spermadhesin genes in porcine male and female reproductive tracts. *Molecular Reproduction & Development,* **61**: 32-41.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D. and Bairoch, A. 2005. *Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook.* J.M. Walker. Totowa, NJ, Humana Press, USA.

Goto, Y., Niwa, Y., Suzuki, T., Dohmae, N., Umezawa, K. and Simizu, S. 2014. C-mannosylation of human hyaluronidase 1: Possible roles for secretion and enzymatic activity. *International Journal of Oncology*, **45**: 344-350.

Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Engineering*, **4**: 155-161.

Hamby, S.E. and Hirst, J.D. 2008. Prediction of glycosylation sites using random forests. *BMC Bioinformatics*, **9**: 1-13.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. and Davies, G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences (USA)*, **92**: 7090-7094.

IKAI, A. 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, **88**: 1895-1898.

Jansen, S., Ekhlasi-Hundrieser, M. and Topfer-Petersen, E. 2001. Sperm adhesion molecules: Structure and function. *Cells Tissues Organs* **168**: 82-92.

Julenius, K. 2007. NetCGlyc 1.0: Prediction of mammalian C-mannosylation sites. *Glycobiology*, **17**: 868-876.

Khan, I.K., Wei, Q., Chitale, M. and Kihara, D. 2015. PFP/ESG: Automated protein function prediction servers enhanced with gene ontology visualization tool. *Bioinformatics*, **31**: 271-272.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, **157**: 105-132.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, **26**: 283-291.

Lathrop, W.F., Myles, C.E. and Primakoff, P. D.G. 1990. cDNA cloning reveals the molecular structure of a sperm surface protein, PH-20, involved in sperm-egg adhesion and the wide distribution of its gene among mammals. *The Journal of Cell Biology*, **111**: 2939-2949.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P. A., Yamashita, R.A., Yin, J.J., Zhang, D. and Bryant, S.H. 2005. CDD: A conserved domain database for protein classification. *Nucleic Acids Research*, **33**(suppl 1): D192-D196.

McLeskey, S.B., Dowds, C., Carballada, R., White, R.R. and Saling, P.M. 1998. Molecules involved in mammalian sperm-egg interaction. *International Review of Cytology*, **177**: 57-113.

Rodrigues, J.P.G.L.M., Levitt, M. and Chopra, G. 2012. KoBaMIN: A knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Research*, **40**(W1): W323-W328.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. 2005. InterProScan: Protein domains identifier. *Nucleic Acids Research*, **33**(suppl 2): W116-W120.

Tung, K.S., Primakoff, P., Woolman-Gamer, L. and Myles, D.G. 1997. Mechanism of infertility in male guinea pigs immunized with sperm PH-20. *Biology of Reproduction*, **56**: 1133-1141.

Wiederstein, M. and Sippl, M.J. 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, **35**(Web Server issue): W407-W410.

Yudin, A.I., Vandevoort, C.A., Li, M.W. and Overstreet, J.W. 1999. PH-20 but not acrosin is involved in sperm penetration of the macaque zona pellucida. *Molecular Reprodroduction & Development*, **53**: 350-362.

Zhang, Y. and Skolnick, J. 2005. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, **33**: 2302-2309.