International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

In silico Screening of Deleterious Missense Variants of Cattle ANPEP Gene Reveals the Impact on Protein Structure and Function

Ymberzal Koul

Sri Venkateshwara Veterinary University, India
Email: *ymberzal[at]gmail.com*

Abstract: ***Alanyl aminopeptidase or aminopeptidase N (ANPEP) gene codes for the digestive enzyme aminopeptidase that plays a role in the final digestion of peptides generated from hydrolysis of proteins by gastric and pancreatic proteases. The ANPEP gene is conserved across many species, thus implying its essentiality and natural selection. A mutation at a conserved sequence causes protein defunctionalisation and is usually not well tolerated. Hence, the present study was conducted to investigate the non - synonymous SNPs of cattle ANPEP gene by using several computational prediction tools. Out of 176 missense variants in ANPEP gene, retrieved from Ensembl database, 54 variants were found to be deleterious based on SIFT score. These variants were subject to further evaluation through PANTHER, PolyPhen2, PredictSNP, SNAP2 and PhD - SNP. Ultimately, out of 54 deleterious nsSNPs, 20 were confirmed as disease - causing through a consensus approach. The impact of mutation on protein stability, studied through I - Mutant 2.0, muPro and mCSM tools, revealed highly destabilizing effect of A411D, F218S, I864S, P79S, Y191S, Y195S variants. The effects of amino acid substitutions on native structure, analysed through Project HOPE server, confirmed that most of the deleterious variants of ANPEP gene are located in the metalloprotease region of the protein, which is highly conserved and essential for normal function and interaction between the domains. Six deleterious nsSNPs revealed from this study require a confirmatory analysis by wet lab studies in a wider population.***

Keywords: ANPEP, mutation, *in silico*, SIFT, cattle, deleterious

1. Introduction

The ANPEP gene, located on chromosome 21 in the bovine genome, encodes aminopeptidase. This enzyme plays a crucial role in the digestion of peptides derived from protein breakdown. It belongs to the zinc - binding metalloproteinase super family, characterized by a pentapeptide consensus sequence in its extracellular carboxyterminal domain. The importance of this gene is well researched in humans. Aminopeptidase N (also known as CD13) is another enzyme in this class and is identical to ANPEP based on sequence comparisons. Aminopeptidase N is involved in the metabolism of regulatory peptides by various cell types, including small intestinal and renal tubular epithelial cells, macrophages, granulocytes, and synaptic membranes in the central nervous system. It acts as a receptor for the HCoV - 229E alphacoronavirus and other non - human coronaviruses. Additionally, this gene has been associated with angiogenesis, tumor growth, and metastasis. Defects in the ANPEP gene are linked to different types of leukemia and lymphoma. In recent years, the ANPEP gene has gained attention in bovine research, particularly in genome - wide association studies (GWAS) related to cattle. Its significance in these studies underscores the need to further investigate its role and impact in bovine biology.

The implementation of large - scale breed improvement programs in dairy herds has resulted in a substantial increase in the number of offspring produced by a limited number of bulls through artificial insemination. However, this practice has inadvertently led to the accumulation of deleterious mutations within the population due to the elevated level of consanguinity. The utilization of a reduced number of sires and the widespread adoption of artificial insemination have contributed to a higher degree of relatedness among individuals within the population. As a result, the chances of inheriting harmful genetic variants from common ancestors have significantly increased. The concentration of deleterious mutations can have detrimental effects on the overall genetic health of the population. These mutations may disrupt vital biological processes, compromise reproductive success, and increase the susceptibility to various hereditary disorders or diseases. Deleterious mutations in the ANPEP gene can have significant consequences on its function and associated biological processes. These mutations can disrupt the enzymatic activity of aminopeptidase, impairing its ability to effectively digest peptides derived from protein breakdown. This can lead to compromised peptide metabolism and dysregulation of cellular processes. Furthermore, deleterious mutations in ANPEP may affect the receptor function of aminopeptidase N, interfering with its interaction with viruses and potentially compromising the host's defense mechanisms against viral infections.

Before the advent of in silico tools, the identification of deleterious mutations relied primarily on experimental methods and observational data. Methods like gene expression studies and functional assays were used to study mutations in a population. Such methods allowed scientists to assess the impact of mutations on gene function and protein structure, but were time consuming and cumbersome. In addition, population studies and pedigree analysis were used to identify associations between specific mutations and disease phenotypes getting inherited in particular families or communities. The emergence of *in silico* tools has revolutionized the field by enabling rapid and efficient prediction of the functional impact of

mutations, facilitating targeted experimental validation, and accelerating our understanding of the role of deleterious mutations.

The present study is targeted towards identifying deleterious mutations in the ANPEP gene of cattle through computational methods. These computational approaches can assess factors such as conservation scores, protein structure, and known functional regions to prioritize and classify mutations as potentially deleterious. By employing these computational methods, the study aims to contribute to our understanding of the genetic diversity and health of cattle populations, shedding light on potential genetic risks and aiding in the development of informed breeding and management strategies.

2. Materials and Methods

1) Dataset collection

The transcripts for the bovine ANPEP gene (ANPEP - 202 and ANPEP - 201) were obtained from the Ensembl database (access date: March, 2023).

2) Identification of deleterious mutations

The tools used included SIFT (Sorting Intolerant From Tolerant), PANTHER (Protein Analysis Through Evolutionary Relationship), PolyPhen - 2 (Polymorphism Phenotyping), PROVEAN (Protein Variation Effect Analyzer), and PredictSNP. To obtain SIFT predictions, we utilized the VEP Web interface, which provided Tolerance Index (TI) scores ranging from 0.0 to 1.0. Mutations with a TI score of 0.05 or less were considered intolerant or deleterious (Ng and Henikoff, 2003). PolyPhen - 2 utilized the PSIC (Position Specific Independent Counts) score to predict the functional effect of deleterious mutations. Predictions were categorized as "probably damaging, " "possibly damaging, " or "benign" (Ramensky et al., 2002). PANTHER estimated the likelihood of non - synonymous single nucleotide polymorphisms (nsSNPs) to cause functional impact on the protein. It calculated the subPSEC (substitution position - specific evolutionary conservation) score and provided predictions of "probably damaging, " "possibly damaging, " or "benign" based on evolutionary conservation (Tang and Thomas, 2016). PROVEAN predicted the effect of amino acid substitutions on protein function using a threshold of - 2.5. A score of ≤ (-) 2.5 was considered deleterious, while a score > (-) 2.5 was considered neutral (Choi et al., 2012). The query protein sequences (in FASTA format) of nsSNPs predicted as deleterious by SIFT were submitted as input to PANTHER, PROVEAN, PolyPhen - 2, and the consensus classifier of PredictSNP server for further analysis.

We employed the consensus classifier of PredictSNP, a computational tool, to obtain predictions from multiple algorithms: MAPP (Multivariate Analysis of Protein Polymorphism), PhD - SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms), and SNAP (screening for non - acceptable polymorphisms). The MAPP predictions were based on evaluating the physicochemical variation in each column of a sequence alignment (Stone and Sidow, 2005). PhD - SNP utilized a support vector machine (SVM) - based classifier to provide predictions as "neutral" or "deleterious" (Capriotti et al., 2006). SNAP, on the other hand, is a neural network - based tool specifically designed to assess the functional effects of single amino acid substitutions in proteins (Bromberg and Rost, 2007). By integrating these diverse prediction methods within PredictSNP, the study aimed to obtain comprehensive insights into the potential deleterious mutations in the ANPEP gene of cattle. Consurf tool was used to assess the conservation score. Conservation scores are a measure of how well - preserved a particular amino acid position is across different species during evolution.

3) Assessment of deleterious mutations on protein stability

Two tools, I - Mutant2.0 and MUpro, use machine learning algorithms to predict changes in protein stability caused by non - synonymous single nucleotide polymorphisms (nsSNPs). They calculate the ΔΔG value (in kcal/mol) for a given mutation and predict whether stability will increase or decrease. These tools consider differences in size and biochemical properties between natural and mutant residues (Capriotti et al., 2005; Cheng et al., 2006). To assess the impact of SNPs on protein structure at a specific amino acid position, the Project HOPE version 1.1.1 tool (https: //www3. cmbi. umcn. nl/hope/) was employed. This tool utilizes various processes, including BLAST against Uniprot and PDB for homology modeling and WHAT IF for tertiary structure data. Additionally, it leverages the Distributed Annotation System (DAS) server to provide additional functionality (Venselaar et al., 2010).

3. Results and Discussion

A total of 174 missense variants were identified from the Ensembl dataset of ANPEP gene. On the basis of SIFT score of variants, 54 out of 174 were found to be deleterious. Upon using all the *in - silico* tools as mentioned in materials and methods, twenty nsSNPs were found as common deleterious (Figure 1).

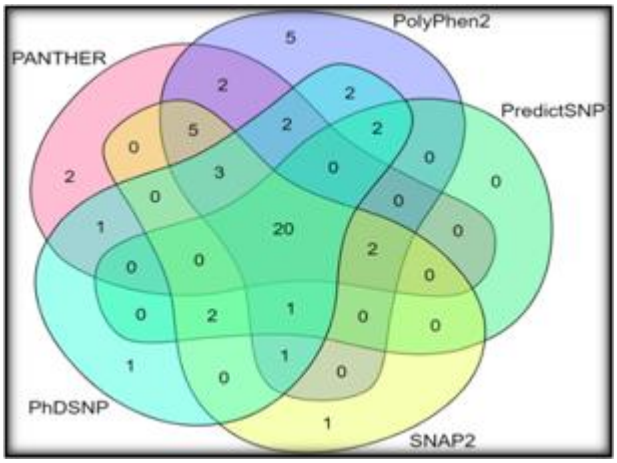

Figure 1: Venn diagram displaying mutations obtained as common from all *in silico* tools

Impact of deleterious mutations on protein stability: The assessment of deleterious mutations on protein stability using the I - Mutant prediction tool revealed that the majority of the variants analyzed in this study led to a decrease in protein stability. However, three specific

variants, S273F, Y853C, and Y890C, were found to have an opposite effect and were predicted to increase protein stability. Using the muPro tool, it was observed that all variants, except for S273F, resulted in a decrease in protein stability. This suggests that these mutations may disrupt the structural integrity of the protein, potentially affecting its proper folding and function. Protein stability is a fundamental property that determines the three - dimensional structure and functional behavior of a protein. It relies on the intricate balance of non - covalent interactions and structural features that contribute to the folded state. Disrupting these interactions and structural elements can lead to misfolding, aggregation, or loss of function.

The mCSM program provided further insights into the impact of specific variants on protein stability. Variants A411D, F218S, I864S, P79S, Y191S, and Y195S were classified as 'highly destabilizing' to the protein structures. This suggests that these mutations significantly disrupt the stability of the protein and may have a detrimental effect on its overall structure and function. On the other hand, the variant D430V was classified as 'stabilizing' to the protein structure, indicating that this mutation might actually enhance the stability of the protein.
Overall, the results indicate that most of the analyzed mutations have a destabilizing effect on the protein structure, potentially compromising its stability and function. However, it's worth noting that specific variants may have differing effects, such as increasing stability (S273F) or stabilizing the protein structure (D430V). These findings contribute to our understanding of the potential impact of these mutations on the ANPEP gene and provide valuable insights for future studies and breeding programs.

Out of the 20 variants analyzed in this study, 7 of them were found to be located on highly conserved positions with a conservation score of 9. A high conservation score of 9 indicates that the amino acid at that specific position is strongly conserved across a wide range of species. This suggests that this particular amino acid residue plays a critical role in maintaining the structure, function, or essential interactions of the protein. The high conservation of these positions implies that any alterations or substitutions at these sites may have a significant impact on the protein's structure or function. Mutations occurring at conserved positions have the potential to disrupt important interactions, affect protein stability, or impair essential functional domains. Identifying variants at highly conserved positions is particularly noteworthy because they are more likely to have functional consequences compared to mutations occurring at less conserved regions. This indicates that these specific positions are crucial for the proper functioning and integrity of the protein. Further investigation and experimental validation of these variants are warranted to elucidate their functional implications and assess their potential role in disease susceptibility or other biological processes associated with the ANPEP gene.

4. Conclusion

In conclusion, this study utilized computational methods to identify deleterious mutations in the ANPEP gene of cattle. The results of the study indicated that most of the analyzed variants had a detrimental effect on protein stability, potentially compromising the proper folding and function of the ANPEP protein. However, a few specific variants were found to have differing effects, such as increasing stability or stabilizing the protein structure. Additionally, it was observed that a subset of variants occurred at highly conserved positions, suggesting their potential functional importance. These variants may significantly disrupt critical protein interactions and affect the overall structure and function of the ANPEP protein.

5. Future Scope

The predicted deleterious mutations should be experimentally validated to confirm their functional impact on the ANPEP protein. Laboratory - based assays, such as site - directed mutagenesis, protein expression and purification, and functional assays, can be performed to assess the effects of these mutations on protein structure, enzymatic activity, protein - protein interactions, and other relevant functional properties. The knowledge gained from this study can be applied to inform breeding strategies aimed at minimizing the negative impact of deleterious mutations on cattle populations. Implementing genomic selection approaches that consider the identified mutations as well as incorporating genetic screening and management practices to avoid mating between carriers of these mutations can help maintain genetic health and improve desired traits in cattle populations.

References

[1] Bromberg, Y., andRost, B. (2007). SNAP: predict effect of non - synonymous polymorphisms on function. Nucleic acids research, 35 (11), 3823 - 3835.
[2] Capriotti, E., Calabrese, R., and Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 22 (22), 2729 - 2734.
[3] Capriotti, E., Fariselli, P., and Casadio, R. (2005). I - Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research, 33 (SUPPL.2), W306–W310.
[4] Cheng, J., Randall, A., and Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. Proteins: Structure, Function, and Bioinformatics, 62 (4), 1125 - 1132.
[5] Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels.
[6] Ng, P. C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research, 31 (13), 3812 - 3814.
[7] Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. Nucleic acids research, 30 (17), 3894 - 3900.
[8] Stone, E. A., &Sidow, A. (2005). Physicochemical constraint violation by missense substitutions

mediates impairment of protein function and disease severity. Genome research, 15 (7), 978 - 986.

[9] Tang, H., and Thomas, P. D. (2016). PANTHER - PSEP: predicting disease - causing genetic variants using position - specific evolutionary preservation. Bioinformatics, 32 (14), 2230 - 2232.

[10] Venselaar, H., te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e - Science approach with life scientist friendly interfaces. BMC Bioinformatics, 11 (1), 548.