



A comprehensive *in silico* prediction of the most deleterious missense variants in the bovine *LEP* gene

MOHAMMED BAQUR S. AL-SHUHAIB *

Department of Animal Production, College of Agriculture, Al-Qasim Green University, Al-Qasim, Iraq

Abstract

Leptin is a versatile hormone involved in many biological functions, including controlling body weight, energy homeostasis, reproduction, and immune function. Though exhaustive studies were performed on the bovine *LEP* gene, no efforts have been made to comprehensively and systematically analyze single nucleotide polymorphisms (SNPs) in its coding sequence. The present study was conducted to identify the most deleterious nonsynonymous SNPs (nsSNPs) of the bovine *LEP* gene. SNPs retrieved from the *dbSNP* database were investigated using various computational tools, including SIFT, PolyPhen-2, PANTHER, PROVEAN, SNAP2, I-Mutant2, mCSM, SDM, DUET, Cobalt, SPPIDER, ConSurf, and MutPred. A total of 28 nsSNPs were considered for the present study. Only 4 nsSNPs, namely, R66M, D186G, C191S, and C191G were found to be deleterious by all the used common nsSNP prediction tools and affected leptin protein structure, function, and biological stability. These variants were located in very highly conserved positions, and thus mutations in these amino acid positions have deleterious evolutionary consequences. The findings of the present study proved that R66M, D186G, C191S, and C191G nsSNPs have the most deleterious consequences on both the structure and the function of bovine leptin, with a special emphasis on the remarkable effects of the last two nsSNPs on the breakage of the disulfide linkage which may lead to a variety of deleterious consequences of this disturbed three-dimensional structure on bovine life and performance. This study provides the first comprehensive computation of the damaging effects of nsSNPs on leptin in bovines.

Key words: cattle, computation, leptin, non-synonymous SNPs

Introduction

One of the most prominent hormones that provides a hotspot indication for numerous metabolic traits of cattle is leptin. Leptin is a hormone that participates in the regulation of body weight by maintaining a balance between food intake and energy expenditure by informing the brain to change the stored levels of energy (Zhou et al., 2009). Leptin plays a vital role in the regulation of feed intake and energy metabolism in cattle (Liefers et al., 2005). In addition to its remarkable role in controlling appetite (Yadav et al., 2011), leptin has other noticeable roles in regulating reproduction, growth, body composition, and immunity (Saleem et al., 2015). Leptin is encoded by the *LEP* gene, which is of about 20 kb. It consists of three exons separated by two

introns, and the exonic portion of the *LEP* gene covers about 15 kb of the bovine genome. Actually, the first exon is truncated in the mature blood circulating hormone. Meanwhile, the other two exons produce fully mature 167 residues by excising the first 24 signal-amino acid residues to produce 16 kDa of blood circulating leptin (Liefers, 2004). Leptin contains a distinctive three-dimensional (3D) four- α -helix bundle structure of A-B-C-D pattern (Kline et al., 1997). This structure is arranged in four sequentially similar, antiparallel, left-hand twisted α -helices bundles that are connected by two cross-over links, alongside one short loop (Gutierrez et al., 2009). In addition to the four main helices, an extra fragment known as helix E, which is a distorted short helix, is also present in the structure. Helix E is found in the

* Corresponding author: Department of Animal Production, College of Agriculture, Al-Qasim Green University, 8-Al-Qasim, Babil 51001, Iraq; e-mail: baquralhilly_79@yahoo.com

loop linking helices C and D. Moreover, leptin has one disulfide bond that connects two cysteine residues (Cys141 and Cys191) within the C and D helices to form a unique kink. This extending loop has been proven to be very crucial for the structural stability and integrity and, therefore, the biological activity of leptin (Haglund et al., 2012). Hence, any missense mutation that changes this highly organized 3D structure of leptin may have deleterious outcomes on many critical metabolic pathways this hormone is involved in. It was recently reported that many missense mutations have several effects on the corresponding proteins due to a single amino acid substitution (Borgio et al., 2016; Al-Shuhaib et al., 2017) or due to the substitution of the transcription factors (Liao and Lee, 2010; Abdulazeez et al., 2019), or due to other reasons (Prokop et al., 2012). It is widely acknowledged that silent mutations have a high potential to cause alterations in protein expression, conformation, and function (Sauna and Kimchi-Sarfaty, 2011). However, mutations within introns and other non-coding loci do not substitute amino acid sequences. Conversely, a nonsynonymous single nucleotide polymorphism (nsSNP), which is present within the exon of a gene, is responsible for the incorporation of an alternative amino acid and known to be one of the main causes of the possible alterations in the mode of action of proteins. However, each amino acid alteration has its own consequences regarding its position and identity in the 3D structure of leptin. Accordingly, it is important to differentiate these consequences computationally (Al-Shuhaib et al., 2018). Noteworthy, several missense mutations have been discovered in this highly studied leptin protein (Liefers et al., 2004; Matteis et al., 2012); however, no comprehensive study has been presented to predict the final consequences of whole nsSNPs in leptin. Therefore, this study was designed to provide the first computational prediction of the most deleterious missense mutations in bovine leptin.

Materials and methods

Dataset used for SNP annotation

Bovine (*Bos taurus*) leptin sequence with National Center of Biotechnology Information (NCBI) accession number XP_010802755.1 was the input FASTA sequence for the present study. The *dbSNP* database was used to retrieve the nsSNP *LEP* gene for this study (<https://www.ncbi.nlm.nih.gov/projects/SNP/>).

Table 1. Distribution of the retrieved SNPs of the bovine *LEP* gene. SNPs in boldface are the non-synonymous SNPs that were selected for the present comprehensive study

Type of SNP	Type of SNP	Number of SNPs	Approximate percentage of SNPs [%]
1	near 5' end SNPs	73	7.2
2	5' UTR SNPs	3	0.3
3	nsSNPs	28	2.7
4	sSNPs	13	1.3
5	intronic SNPs	646	63.3
6	3' UTR SNPs	239	23.4
7	near 3' end SNPs	18	1.8
Total		1020	100

The retrieved SNPs were classified into several types as mentioned in (Table 1). The current study was focused only on the analysis of the SNPs that substitute amino acids sequences, nsSNPs or amino acid variants. A total of 28 nsSNPs distributed over 3 exons of the *LEP* gene were retrieved (Table 2). Subsequently, a variety of state-of-the-art *in silico* tools were used to assess the consequences of these nsSNPs on the structure, function, and stability of leptin.

Finding deleterious nsSNPs by SIFT, PANTHER, Polyphen-2, PROVEAN, and SNAP2

The structural and functional consequences of all retrieved nsSNPs in the bovine leptin protein were analyzed by several *in silico* tools, including SIFT (Sorting Intolerant from Tolerant SNPs) (Ng and Henikoff, 2003), PANTHER (Protein Analysis Through Evolutionary Relationship) (Tang and Thomas, 2016), Polyphen-2 (Polymorphism Phenotyping v2) (Adzhubei et al., 2013), PROVEAN (Protein Variation Effect Analyzer), and SNAP2 (Smigielski et al., 2000). Substitutions predicted by SIFT that had a tolerance index of less than 0.05 were predicted to be "intolerant" or "deleterious"; those greater than or equal to 0.05 were predicted to be "tolerated" (Ng and Henikoff, 2003). The expected PANTHER score of each nsSNP was based on a threshold of 450 my, or estimated preservation time. Thus, an nsSNP was classified as "probably damaging" (time > 450 my), 'possibly damaging' (450 my > time > 200 my), and "probably benign" (time < 200 my) (Tang and Thomas, 2016). Three common scores were obtained from Polyphen-2, "pro-

Table 2. The position and distribution of nsSNPs in bovine leptin protein

Number	Chromosome number 4 position	Exonic position	dbSNP ID	nsSNP
1	93249846	exon No. 1	rs448185719	M1L
2	93249850	exon No. 1	rs460213906	E2V
3	93249852	exon No. 1	rs478688866	P3A
4	93249856	exon No. 1	rs445876293	R4L
5	93249858	exon No. 1	rs464306478	R5G
6	93249867	exon No. 1	rs477404171	G8R
7	93249871	exon No. 1	rs469453331	I9N
8	93249889	exon No. 1	rs455111977	A15G
9	93262003	exon No. 2	rs29004487	Y32F
10	93262056	exon No. 2	rs29004488	C50R
11	93262101	exon No. 2	rs436985239	T65A
12	93262105	exon No. 2	rs482037251	R66M
13	93263939	exon No. 3	rs475359339	H91N
14	93263940	exon No. 3	rs442451048	H91P
15	93263979	exon No. 3	rs29004508	A104V
16	93264009	exon No. 3	rs525274306	P114H
17	93264036	exon No. 3	rs380222501	N123S
18	93264107	exon No. 3	rs479601164	R147G
19	93264127	exon No. 3	rs440180195	E153D
20	93264152	exon No. 3	rs469111163	S162P
21	93264169	exon No. 3	rs518320795	E167D
22	93264205	exon No. 3	rs434902010	Q179H
23	93264207	exon No. 3	rs453354732	D180A
24	93264224	exon No. 3	rs465539127	D186Y
25	93264225	exon No. 3	rs432552158	D186G
26	93264233	exon No. 3	rs457585073	P189T
27	93264239	exon No. 3	rs442437081	C191S
28	93264239	exon No. 3	rs442437081	C191G

bably damaging”, “possibly damaging”, and “possibly benign” based on the scores that ranged from “0” to “1” (Adzhubei et al., 2013). PROVEAN allowed for a balanced separation between neutral and deleterious amino acids by relying on a threshold of -2.5 , i.e., the nsSNP is “deleterious” when it scores ≤ -2.5 and is “neutral” when it scores > -2.5 (Choi et al., 2012). SNAP2 generated automatic grading that ranged from “damaging” (score $> zero$) to non-damaging “neutral” (score $< zero$) and (Smigielski et al., 2000).

Assessment of nsSNP effect on protein stability by using I-Mutant 2.0, mCSM, SDM, and DUET

To obtain a better assessment of the stability of the bovine leptin protein impacted by missense mutations, the retrieved nsSNPs were analyzed using several tools, including I-Mutant 2.0 (Capriotti et al., 2005), mCSM (Pires et al., 2012), SDM (Worth et al., 2012), and DUET (Pires et al., 2014). In the case of I-Mutant 2.0, the input data were FASTA sequences of the referring leptin protein, whereas the input data in mCSM, SDM,

and DUET were submitted as protein data bank (PDB) files. In all prediction cases, the submitted file of the referring leptin protein was computed along with its amino acid substitutions and analyzed in terms of free energy change ($\Delta\Delta G$) values (kcal/mol). Negative values of $\Delta\Delta G$ were destabilizing, while positive values of $\Delta\Delta G$ were stabilizing to the 3D structure of the altered proteins.

Identifying evolutionary, conserved, and functional regions by using NCBI COBALT, ConSurf, SPPIDER, and MutPred tools

The evolutionary, conserved regions were initially identified using a multiple sequence alignment viewer 1.5.2 and the NCBI COBALT tool (<https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi>). The bovine *LEP* gene was manually compared with sequences from seventeen selected sources of eutherian mammalian DNA, including *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Gorilla gorilla gorilla* (gorilla), *Pongo abelii* (orangutan), *Macaca mulatta* (macaque), *Pan paniscus* (bonobo), *Papio Anubis* (olive baboon), *Ovis aries* (sheep), *Capra hircus* (goat), *Equus caballus* (horse), *Sus scrofa* (pig), *Canis lupus familiaris* (dog), *Felis catus* (cat), *Oryctolagus cuniculus* (rabbit), *Callithrix jacchus* (marmoset), *Rattus norvegicus* (rat), and *Mus musculus* (mouse). In ConSurf tool, both homology-based amino acid sequence alignment and empirical, Bayesian algorithm-based calculations were made to predict the evolutionary status of the position of each amino acid residue (Ashkenazy et al., 2010). SPPIDER, a virtual, web-based server that belongs to the POLYVIEW prediction system (<http://polyview.cchmc.org/polyview3d.html>), was used to compare the conservative evolutionary status of bovine leptin. Subsequently, the specific role of the observed harmful nsSNPs and their potential effect on protein functionality were analyzed using the MutPred tool (Pejaver et al., 2017). In all prediction cases, the referring amino acid sequences of leptin protein were submitted in FASTA format.

3D modeling of the virtual protein structure

Whole amino acid sequences of bovine leptin were retrieved from the NCBI web server. The UniProtKB/Swiss-Prot entry number of this protein is P50595. Since no matching PDB entries for the full 3D structure of leptin were found (<http://www.uniprot.org>), a 3D structure of bovine leptin was generated online by the RaptorX

software (Källberg et al., 2012). The observed substitutions within its corresponding altered proteins were virtually visualized using the PyMol-v1, 7.0.1 tool (www.shrodinger.com).

Prediction of nsSNP participation in ligand binding sites through FTsite tool

The FTsite machine is an online server that was utilized to predict ligand binding sites of bovine leptin (<http://ftsitesite.bu.edu>). It recognizes the binding sites precisely, and provides a structure-based explanation of functional relationships among proteins (Ngan et al., 2012). The query protein sequence data were submitted as RaptorX-built PDB files.

Results

In this study, a series of *in silico* prediction analysis methods were used to analyze all the amino acid substitutions in the coding region of the bovine *LEP* gene (gene ID 280836). Therefore, all the nsSNPs were retrieved from dbSNP database. By using several computational tools and checking their effect on the structure, the function, the stability of bovine leptin as well as the evolutionary conservation scores regarding the corresponding amino acid residues in bovine protein were determined.

Out of a total of 1020 SNPs retrieved for the bovine *LEP* gene, 239 in 3' UTR, 3 near the 5' end of the coding sequence, 646 in introns, 18 near the 3' end of the coding sequence, and 13 synonymous SNPs, and only nsSNPs that represented 3% (28) of the total SNPs, were selected for further analysis (Fig. 1A). After their retrieval, different software tools were used for the structural and functional annotation of these nsSNPs. The used software tools included several state-of-the-art bio-computational tools, including SIFT, PolyPhen, PROVEAN, SNAP2, PANTHER, I-Mutant 2.0, mCSM, SDM, and DUET, which were used to evaluate the potential consequences of the deleterious effects of SNPs on leptin structure, function, and stability. By comparing the cumulative computations of these methods, the degree of severity of the damaging consequences of each particular nsSNP was verified. Though the SIFT tool is commonly used to provide evolutionary data on the impact of each variant on both protein structure and function (Ng and Henikoff, 2006), its insufficiency is attributed to the low availability of stored non-human data in

Table 3. The *in silico* analysis of the observed nonsynonymous SNPs on bovine leptin using several bioinformatics tools. The boldface nsSNPs refer to the totally damaging effect as predicted by PolyPhen-2/PANTHER/PROVEAN/SNAP2/Mutant-2/mCSM/SDM/DUET prediction tools. The nsSNPs are arranged in ascending order, from the least to the most deleterious consequences on the altered protein

nsSNP	SIFT	PolyPhen-2	PANTHER	PROVEAN	SNAP2	Mutant-2	mCSM	SDM	DUET	Score/ Prediction
E2V	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	-36	0.47	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	neutral	increase	unknown	unknown	unknown	prediction
M1L	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	-10	-0.56	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	neutral	decrease	unknown	unknown	unknown	prediction
P3A	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	-6	-0.76	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	neutral	decrease	unknown	unknown	unknown	prediction
R4L	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	2	0.29	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	effect	decrease	unknown	unknown	unknown	prediction
R5G	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	58	-0.73	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	effect	decrease	unknown	unknown	unknown	prediction
G8R	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	73	0.43	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	effect	decrease	unknown	unknown	unknown	prediction
I9N	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	66	-0.44	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	effect	decrease	unknown	unknown	unknown	prediction
A15G	0.00 <small>Low Confidence Prediction</small>	not scored	not scored	0.000	-58	-0.25	not scored	not scored	not scored	score
	affect protein function	unknown	invalid SNP	neutral	neutral	decrease	unknown	unknown	unknown	prediction
Y32F	1.0	0.000	1	-0.278	-54	-0.26	not scored	not scored	not scored	score
	tolerated	benign	benign	neutral	neutral	decrease	unknown	unknown	unknown	prediction
C50R	0.94	0.000	1	2.547	-19	-0.31	0.333	0.52	0.718	score
	tolerated	benign	benign	neutral	neutral	decrease	stabilize	stabilize	stabilize	prediction
D180A	0.08	0.017	176	-3.017	6	0.18	0.162	1.48	0.667	score
	tolerated	benign	benign	deleterious	effect	increase	stabilize	stabilize	stabilize	prediction
R147G	0.76	0.002	97	-1.512	30	-1.63	-0.06	0.47	0.224	score
	tolerated	benign	benign	neutral	effect	decrease	destabilize	stabilize	stabilize	prediction
E153D	0.40	0.005	176	-0.633	7	-0.30	-0.302	-0.41	-1.19	score
	tolerated	benign	benign	neutral	effect	decrease	destabilize	destabilize	destabilize	prediction

T65A	0.08	0.005	220	-1.558	-9	-2.14	-0.654	2.06	0.168	score
	tolerated	benign	damaging	neutral	neutral	decrease	destabilize	stabilize	stabilize	prediction
P114H	0.21	0.064	176	-5.513	13	-1.90	-0.117	-0.39	0.039	score
	tolerated	benign	benign	deleterious	effect	decrease	destabilize	destabilize	stabilize	prediction
D186Y	0.00 <small>Low Confidence Prediction</small>	1.000	220	-6.207	69	0.05	0.164	0.67	0.114	score
	affect protein function	damaging	damaging	deleterious	effect	increase	stabilize	stabilize	stabilize	prediction
N123S	0.02 <small>Low Confidence Prediction</small>	0.798	220	-3.134	-9	-1.12	-0.27	0.43	0.045	score
	affect protein function	damaging	damaging	deleterious	neutral	decrease	destabilize	stabilize	stabilize	prediction
S162P	0.02 <small>Low Confidence Prediction</small>	0.998	220	-3.265	82	0.25	-0.214	0.3	0.09	score
	affect protein function	damaging	damaging	deleterious	effect	increase	destabilize	stabilize	stabilize	prediction
Q179H	0.02 <small>Low Confidence Prediction</small>	0.044	176	-2.780	-33	-0.10	-0.868	-0.19	-0.628	score
	affect protein function	benign	benign	deleterious	neutral	decrease	destabilize	destabilize	destabilize	prediction
P189T	0.03 <small>Low Confidence Prediction</small>	0.519	176	-2.408	-28	-1.36	-1.079	-0.45	-0.852	score
	affect protein function	damaging	benign	neutral	neutral	decrease	destabilize	destabilize	destabilize	prediction
H91P	0.00 <small>Low Confidence Prediction</small>	0.928	220	-4.509	27	-0.58	0.35	-1.35	0.099	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	stabilize	destabilize	stabilize	prediction
E167D	0.00 <small>Low Confidence Prediction</small>	0.995	220	-2.062	69	-0.32	-0.636	-1.4	-0.493	score
	affect protein function	damaging	damaging	neutral	effect	decrease	destabilize	destabilize	destabilize	prediction
H91N	0.01 <small>Low Confidence Prediction</small>	0.646	220	-3.038	9	-3.02	-0.499	-0.22	-0.472	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction
A104V	0.00 <small>Low Confidence Prediction</small>	0.600	220	-2.720	4	-0.55	-0.739	-1.69	-0.735	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction
R66M	0.00 <small>Low Confidence Prediction</small>	1.000	220	-4.987	73	-0.66	-0.736	-0.62	-0.772	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction
D186G	0.00 <small>Low Confidence Prediction</small>	0.998	220	-4.971	61	-0.44	-0.585	-0.24	-0.307	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction
C191S	0.00 <small>Low Confidence Prediction</small>	1.000	220	-7.880	73	0.26	-0.785	-0.45	-0.467	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction
C191G	0.00 <small>Low Confidence Prediction</small>	1.000	220	-9.232	81	-0.49	-1.581	-1.09	-1.595	score
	affect protein function	damaging	damaging	deleterious	effect	decrease	destabilize	destabilize	destabilize	prediction

The scores of PANTHER were taken as preservation time. The scores of Mutant2, mCSM, SDM, and DUET were taken as free energy change (DDG)

web-based bio-information servers. Therefore, SIFT predictions were compensated by other comparable prediction tools such as PolyPhen-2, PANTHER, PROVEAN, and SNAP2 through analyzing the effect of these nsSNPs on leptin function. These confirmations came from the ability of PolyPhen-2, PANTHER, PROVEAN, and SNAP2 tools to give compatible results regarding the collective effects of these 28 nsSNPs. Accordingly, eight nsSNPs, M1L, E2V, P3A, R4L, R5G, G8R, I9N, and A15G were entirely eliminated from further predictions since most prediction tools did not give reliable consequences regarding these missense mutations (Table 3).

The main reason behind the elimination of M1L, E2V, P3A, R4L, R5G, G8R, I9N, and A15G nsSNPs from further analyses may be a consequence of their positions in the premature leptin structure. In other words, most of these benign nsSNPs were positioned in the first 24 amino acids which are cleaved-off from the structure of the premature leptin before it is circulated into blood (Liefers, 2004). The ranges of the deleterious consequences of the remaining 20 nsSNPs were further categorized according to the impact of their effect on leptin biological structure and function. Out of the identified 20 nsSNPs, 9 nsSNPs identified in 6 amino acids positions, R66M, H91N, H91P, A104V, S162P, D186Y, D186G, C191S, and C191G were found to represent the most dangerous missense mutations observed in bovine leptin (Fig. 1B). These variants were re-predicted as deleterious by all the four, PolyPhen-2, PANTHER, PROVEAN, and SNAP2 programs and are likely to affect the structure and function of the protein, resulting in different metabolic problems. In all 9 nsSNPs were predicted as deleterious by all the software tools used (R66M, H91N, H91P, A104V, S162P, D186G, D186Y, C191S, and C191G) (Figure 1C). To evaluate the effect of these deleterious nsSNPs on protein stability upon mutation, four *in silico* tools were utilized, i.e. I-Mutant 2, mCSM, SDM, and DUET (Capriotti et al., 2005; Pires et al., 2012; Worth et al., 2012; Pires et al., 2014). These tools further eliminated three additional deleterious nsSNPs as they increased leptin stability upon mutation (H91P, S162P, and D186Y) (Fig. 1D). Thus, by using the deleterious prediction filter and the stability prediction filter, only 6 nsSNPs were found to be deleterious by all these tools, namely R66M, H91N, A104V, D186G, C191S, and C191G. Subsequently, a third layer of pre-

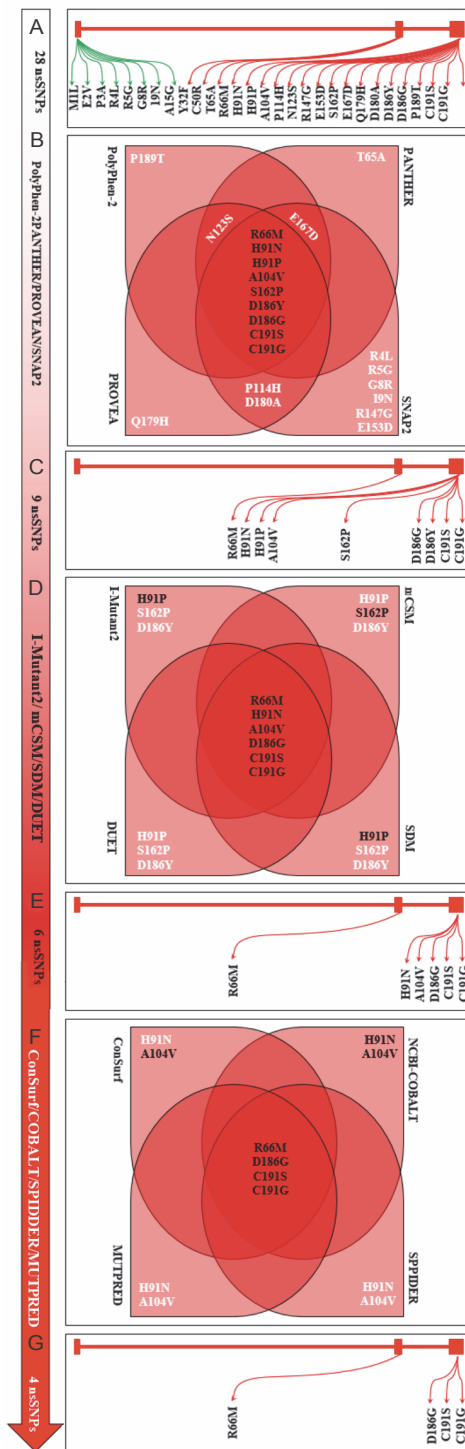


Fig. 1. A schematic diagram of mutation prediction tools showing the most deleterious nsSNPs of bovine leptin. A) out of the 29 retrieved nsSNPs, B) only 9 were found to be entirely deleterious by all tools (PolyPhen-2/PANTHER/PROVEAN/ SNAP2), C) out of the 9 deleterious nsSNPs, D) only 6 were found entirely destabilizing by Mutant2/mCSM/SDM/DUET tools, E) out of the 6 deleterious nsSNPs, F) only 4 were found entirely destabilizing by the COBALT/ConSurf/SPIDDER/MutPred tools, G) the filtered-out entirely deleterious and destabilizing nsSNPs were only four

Table 4. MutPred predictions for the most deleterious nsSNPs observed in the bovine LEP gene

Substitution	Molecular mechanisms with P -values ≤ 0.05	Probability	P -value
R66M	MutPred score 0.830		
	loss of allosteric site at R66	0.46	1.9e-04
	altered transmembrane protein	0.29	2.1e-04
	altered disordered interface	0.28	0.04
	altered ordered interface	0.27	0.05
	altered DNA binding	0.24	0.01
D186G	MutPred score 0.536		
	gain of loop	0.27	0.03
	altered transmembrane protein	0.23	2.0e-03
C191S	MutPred score 0.721		
	gain of intrinsic disorder	0.50	1.7e-03
	gain of B-factor	0.32	1.2e-03
	altered transmembrane protein	0.31	1.2e-04
	loss of disulfide linkage at C191	0.10	0.05
C191G	MutPred score 0.721		
	gain of intrinsic disorder	0.39	0.01
	gain of B-factor	0.35	2.2e-04
	altered transmembrane protein	0.29	4.0e-04
	gain of loop	0.27	0.03
	loss of disulfide linkage at C191	0.10	0.05

diction filters was added to get a further confirmation on the observed variants in terms of evolutionary conservation grade they had and the mechanism of their effect. This prediction task was performed by four prediction tools, NCBI-Cobalt, ConSurf, SPPIDER, and Mutpred. Using NCBI-Cobalt, all 6 nsSNPs were found to be conserved in comparison with all other eutherian animals. The ConSurf tool provided an accurate conservation analysis in which the majority of nsSNPs (R66M, D186G, C191S, and C191G) ranged between conserved and highly conserved with the exception of the 91st amino acid position that consisted of two different variants. Moreover, results obtained with the SPPIDER tool concerning R66M, H91N, H91P, D186G, C191S, and C191G were confirmed by those from ConSurf (Supplementary Fig. 1A, 1B, 1C). The most conserved nsSNPs among the results obtained with (all three) NCBI-Cobalt, ConSurf, and SPPIDER tools were only four (R66M, D186G, C191S, and C191G) (Fig. 1E). Ad-

ditionally, more predictions were offered by ConSurf and SPPIDER to provide further prioritization amongst these 4 conserved variants. Both ConSurf and SPPIDER tools gave similar “very high” conservation for R66M, C191S, and C191G nsSNPs.

To go beyond this point, the ConSurf tool unmasked many more details concerning these three nsSNPs. It revealed that both, the 66th and the 191th amino acids were positioned in functional places within the corresponding exposed positions in bovine leptin. Furthermore, several mechanisms were suggested by MutPred for the observed highly deleterious nsSNPs, such as loss of the allosteric site, altered transmembrane, and disordered DNA-binding properties in R66M, gain of loop and altered transmembrane properties in D186G, and gain of loop and intrinsic disorder, altered transmembrane protein, and loss of disulfide linkage in both C191S and C191G (Table 4).

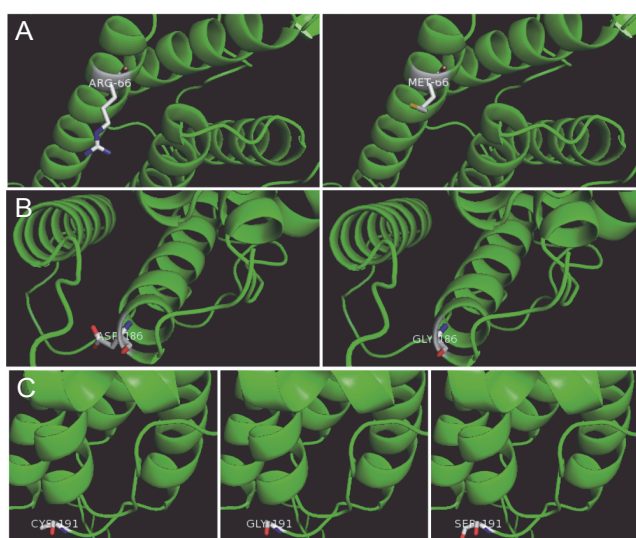


Fig. 2. The postulated 3D structure of the most deleterious missense SNPs in bovine leptin before and after mutation. A–C) branches highlight the involved amino acids in their wild types and mutant forms, respectively

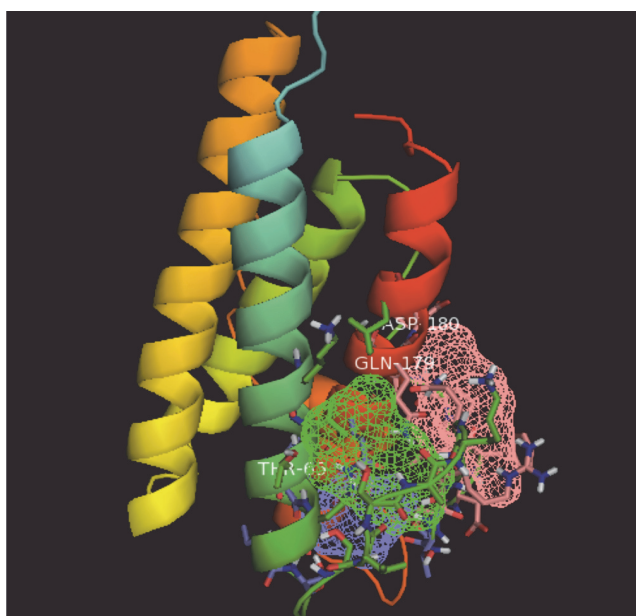


Fig. 3. FTsite prediction of the 3D constructed bovine leptin

All of the constructed 191 (100%) amino acid residues were modeled successfully and only 14 (28%) positions were predicted as disordered by the RaptorX modeling tool. Secondary structures of the native leptin model revealed 53% helix, 1% beta sheet, and 45% loop structures. Although the R66M, C191S, and C191G variants exerted high critical functionality in all predic-

tion tools, the analysis of the 3D putative model of bovine leptin did not predict any involvement of these three variants, as long as all other damaging nsSNPs, in any ligand bound with the leptin receptor was as suggested by FTsite search tool (Fig. 3). Thus, the present study found that all these proven deleterious variants do not participate in the binding with the receptor (supplementary Table 1).

Discussion

There were several reasons that prompted us to engage in the study of the influence of SNPs on the *LEP* gene sequence and their potential influence on protein function (Raschia et al., 2018). They were: 1) many nsSNPs have been reported in the coding region of the *LEP* gene that has a relation to growth, nutritional status, and body composition (Liao et al., 2010), 2) *LEP* gene is located within the quantitative trait locus on chromosome “4” that is directly or indirectly related to several productive traits (Agarwal et al., 2009), 3) the *LEP* gene is a key factor for getting a deeper insight into its association with the leptin receptor in cattle (Trakovická et al., 2013), 4) the combination of the pattern of the *LEP* gene binding with its receptor and its location on chromosome 4 makes the *LEP* gene a brilliant candidate gene that may participate interactively in animal genetic breeding strategies (Jecminkova et al., 2018), 5) the study of nsSNPs of *LEP* gene might have a remarkable consequence on its corresponding protein structure, expression, and function. Accordingly, the most deleterious missense mutations of 191 amino acids of the leptin protein were comprehensively covered in this study.

3D analyses of bovine leptin have identified three functionally important receptor binding sites on the four-helix leptin structure (Zhang et al., 2005). Leptin receptor binding site-I is located in the C-terminus (Peelman et al., 2004). It is a 50-amino-acid long chain, positioned within 141–191 residues in bovine leptin. The C-terminus exhibits a unique structure which might allow leptin to specifically bind to its receptor by enhancing the activity of the N-terminus (Peelman et al., 2014). This is due to the binding of the C-terminus with the N-terminus with the Cys141-Cys191 disulfide bond (Denver et al., 2011). Therefore, any amino acid substitution at this critical site would cause a misfolding in the N-terminus por-

tion and might break its interconnection with the C-terminus. Thus, both C191S and C191G nsSNPs would prevent the N-terminal portion from maintaining the correct leptin conformation needed for binding to its receptor. Thus, C191S and C191G may have deleterious consequences on leptin activity. The Mutpred tool indicated that both C191S and C191G disrupted the Cys141-Cys191 disulfide connection and were involved in this damaging mechanism. With regard to receptor binding site-II, which is positioned within bovine leptin amino acid residues 131–165 (Hiroike et al., 2000), no highly deleterious nsSNP was detected at this site. On the other hand, three completely deleterious nsSNPs (R66M, A104V, H91N, and H91P) were detected at the leptin receptor binding site-III. This binding site is located around the N-terminus of leptin (Iserentant et al., 2005). It extends from residues 47–140 of the mature bovine leptin (Fischer, 2008).

Thus, the presence of R66M, A104V, H91N, and H91P snSNPs in mature leptin might cause improper folding of the N-terminus. Therefore, the present findings indicate that C191S, C191G, A104V, and R66M have the most dangerous consequences due to their highly deleterious impacts predicted by all utilized *in silico* tools. As determined by the ConSurf tool in the present study, these highly damaging amino acids are located in highly evolutionary, conserved positions in bovine leptin. Eventually, the obtained perspective of bovine leptin is the result of a cumulative combination of various computational tools that have collectively provided these data.

Conclusions

This manuscript presents pilot data that filter out the probability of the nsSNPs altering leptin structure and its biological activity. These SNPs may potentially have a profound influence on the protein structure and its ability to interact with receptors. This study detected four amino acid variants, C191S, C191G, A104V, and R66M that, according to the prediction tools used in the study, had the most damaging effects on the protein in terms of structure, function, and stability. R66M and A104V exert their effect primarily by causing loss of the allosteric site and loop alteration, respectively. Importantly, both C191S and C191G directly disrupt disulfide linkage formation in the protein. The last disulfide breakdown is

a natural outcome for the altered protein due to the positioning of both C191S and C191C in cys141–cys191 in the disulfide linkage site of leptin. Therefore, these *in silico* prediction data may provide the first bio-computational evidence of the consequences of the most deleterious missense mutations in bovine leptin in terms of its structure, function, and stability.

The present study provides a starting point for the future targeted breeding-based studies that may be useful for establishing the potential damaging impact of amino acid substitutions on the biological function of leptin in the marker-assisted selection of cattle. These computational findings would/may/might (also) make breeders rethink current breeding systems and may establish *in silico* approaches to them.

References

- Adzhubei I., Jordan D.M., Sunyaev S.R. (2013) *Predicting functional effect of human missense mutations using PolyPhen-2*. Chapter 7: Unit 7.20. Curr. Protoc. Hum. Genet.
- Abdulazez S., Sultana S., Almandil N.B., Almohazey D., Bency B.J., Borgio J.F. (2019) *The rs61742690 (S783N) single nucleotide polymorphism is a suitable target for disrupting BCL11A-mediated foetal-to-adult globin switching*. PLoS ONE 14(2): e0212492.
- Agarwal R., Rout P.K., Singh S.K. (2009) *Leptin: a biomolecule for enhancing livestock productivity*. Indian J. Biotechnol. 8: 169–176.
- Al-Shuhaib M.B.S., Al-Fihan R.A., Al-Qutbi A.A., Al-Thuwaini T.M. (2017) *Potential consequences of DGAT2 and BTN genes polymorphism in Iraqi Holstein cattle*. SAB 4: 127–141.
- Al-Shuhaib M.B.S., Al-Kafajy F.R., Badi M.A., Abdulazez S., Marimuthu K., Al-Juhaishi H.A.I., Borgio J.F. (2018) *Highly deleterious variations in COX1, CYTB, SCG5, FK2, PRL and PGF genes are the potential adaptation of the immigrated African ostrich population*. Comput. Biol. Med. 100: 17–26.
- Ashkenazy H., Erez E., Martz E., Pupko T., Ben-Tal N. (2010) *ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids*. Nucl. Acids Res. 38: W529–W533.
- Borgio J.F., Al-Madan M.S., Abdulazez S. (2016) *Mutation near the binding interfaces at alpha-hemoglobin stabilizing protein is highly pathogenic*. Am. J. Transl. Res. 8(10): 4224–4232.
- Capriotti E., Fariselli P., Casadio R. (2005) *I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure*. Nucl. Acids Res. 33: W306–W310.
- Choi Y., Sims G.E., Murphy S., Miller J.R., Chan A.P. (2012) *Predicting the functional effect of amino acid substitutions and indels*. PLoS ONE 7: e46688.

- Denver R.J., Bonett R.M., Boorse G.C. (2011) *Evolution of leptin structure and function*. Neuroendocrinology 94: 21–38.
- Fischer J.D.H. (2008) *Distributions and associations of single nucleotide polymorphisms in the leptin gene of Bos taurus and Bos indicus cattle*. LSU Master's Thesis: 4301.
- Gutierrez D.A., Puglisi M.J., Hasty A.H. (2009) *Impact of increased adipose tissue mass on inflammation, insulin resistance, and dyslipidemia*. Curr. Diabetes Rep. 9: 26–32.
- Haglund E., Sułkowska J.I., He Z., Feng G-S, Jennings P.A., Onuchic J.N. (2012) *The unique cysteine knot regulates the pleiotropic hormone leptin*. PLoS ONE 7 e45654.
- Hiroike T., Higo J., Jingami H., Toh H. (2000) *Homology modeling of human Leptin/Leptin receptor complex*. Biochem. Biophys. Res. Co. 275: 154–158.
- Jecminkova K., Müller U., Kyselova J., Sztankoova Z., Zavadilova L., Stipkova M., Majzlik I. (2018) *Association of leptin, toll-like receptor 4, and chemokine receptor of interleukin 8 C-X-C motif single nucleotide polymorphisms with fertility traits in Czech Fleckvieh cattle*. Asian-Australas. J. Anim. Sci. 31(11): 1721–1728.
- Källberg M., Wang H., Wang S., Peng J., Wang Z., Lu H., Xu J. (2012) *Template-based protein structure modeling using the RaptorX web server*. Nat. Protoc. 19: 1511–1522.
- Kline A.D., Becker G.W., Churgay L.M., Landen B.E., Martin D.K., Muth WL, Rathnachalam R., Richardson J.M., Schoner B., Ulmer M., Hale J.E. (1997) *Leptin is a four-helix bundle: Secondary structure by NMR*. FEBS Lett. 407: 239–242.
- Liao P.Y., Lee K.H. (2010) *From SNPs to functional polymorphism: the insight into biotechnology applications*. Biochem. Eng. J. 49: 149–158.
- Liefers S. (2004) *Physiology and genetics of leptin in per-parturient dairy cows*. Wageningen University. Ph. D. Thesis.
- Liefers S.C., Veerkamp R.F., TePas M.F.W., Delavaud C., Chilliard Y., Platjie M., Lende V.D. (2005) *Leptin promoter mutations affect leptin levels and performance traits in dairy cows*. Anim. Genet. 36: 111–118.
- Matteis G., Scatà M., Grandoni F., Petrera F., Abeni F., Cattillo G., Napolitano B., Moiola B. (2012) *Association analyses of single nucleotide polymorphisms in the leptin and leptin receptor genes on milk and morphological traits in Holstein cows*. OJAS 2: 174–182.
- Ng P.C., Henikoff S.H. (2003) *SIFT: predicting amino acid changes that affect protein function*. Nucl. Acids Res. 31: 3812–3814.
- Ng P.C., Henikoff S. (2006) *Predicting the effects of amino acid substitutions on protein function*. Annu. Rev. Genom. Hum. Genet. 7: 61–80.
- Ngan C.H., Hall D.R., Zerbe B., Grove L.E., Kozakov D., Vajda S. (2012) *FTSite: high accuracy detection of ligand binding sites on unbound protein structures*. Bioinformatics 15: 286–287.
- Peelman F., VanBeneden K., Zabeau L., Iserentant H., Ulrichts P., Defea D., Verhee A., Catteeuw D., Elewaut D., Tavemier J. (2004) *Mapping of the leptin binding sites and design of a leptin antagonist*. J. Biol. Chem. 279: 41038–41046.
- Peelman F., Zabeau L., Mobarana K., Savvides S.N., Tavemier J. (2014) *20 years of leptin: insights into signaling assemblies of the leptin receptor*. J. Endocrinol. 223(1): T9-23.
- Pejaver V., Urresti J., Lugo-Martinez J., Pagel K.A., Lin G.N., Nam H., Mort M., Cooper D.N., Sebat J., Iakoucheva L.M., Mooney S.D., Radivojac P. (2017) *MutPred2: inferring the molecular and phenotypic impact of amino acid variants*. bioRxiv 134981.
- Pires D.E.V., Ascher D.B., Blundell T.L. (2012) *mCSM: predicting the effects of mutations in proteins using graph-based signatures*. Bioinformatics 30: 335–342.
- Pires D.E., Ascher D.B., Blundell T.L. (2014) *DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach*. Nucl. Acids Res. 42: W314–W319.
- Prokop J.W., Duff R.J., Ball H.C., Copeland D.L., Londraville R.L. (2012) *Leptin and leptin receptor: analysis of a structure to function relationship in interaction and evolution from humans to fish*. Peptides 38(2): 326–336.
- Raschia M.A., Nani J.P., Maizon D.O., Beribe M.J., Amadio A.F., Poli M.A. (2018) *Single nucleotide polymorphisms in candidate genes associated with milk yield in Argentinean Holstein and Holstein x Jersey cows*. J. Anim. Sci. Technol. 12(60): 31.
- Sauna Z.E., Kimchi-Sarfaty C. (2011) *Understanding the contribution of synonymous mutations to human disease*. Nat. Rev. Genet. 12: 638–691.
- Smigielski E.M., Sirotkin K., Ward M., Sherry S.T. (2000) *dbSNP: a database of single nucleotide polymorphisms*. Nucl. Acids Res. 28: 52–355.
- Saleem A.H., Hussain T., Tahir M.Z., Ali A., Khan W.A., Dawood M., et al. (2015) *Role of leptin in growth, reproduction and milk production in farm animals – a review*. JAVAR 3(5): 302–307.
- Tang H., Thomas P.D. (2016) *PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation*. Bioinformatics 32: 2230–2232.
- Trakovická A., Moravčíková N., Kasarda R. (2013) *Genetic polymorphisms of leptin and leptin receptor genes in relation with production and reproduction traits in cattle*. Acta Biochim. Pol. 60: 783–787.
- Worth C.L., Preissner R., Blundell T.L. (2012) *SDM – a server for predicting effects of mutations on protein stability and malfunction*. Nucl. Acids Res. 39: W215–W222.
- Yadav V.K., Oury F., Tanaka K.F., Thomas T., Wang Y., Cremers S., Hen R., Krust A., Chambon P., Karsenty G. (2011) *Leptin-dependent serotonin control of appetite: temporal specificity, transcriptional regulation, and therapeutic implications*. J. Exp. Med. 208(1): 41–52.
- Zhang F., Chen Y., Heiman M., Dimarchi R. (2005) *Leptin: structure, function and biology*. Vitam. Horm. 71: 345–72.
- Zhou H., Hickford J.G., Gong H. (2009) *Identification of allelic polymorphism in the ovine leptin gene*. Mol. Biotechnol. 41: 22–25.