







***In silico* prediction of the functional and structural consequences of the non-synonymous single nucleotide polymorphism A122V in bovine CXC chemokine receptor type 1**

A. F. Guzzi^a , F. S. L. Oliveira^a , M. M. S. Amaro^a ,
P. F. Tavares-Filho^a and J. E. Gabriel^{a*} 

^aUniversidade Federal do Vale do São Francisco – UNIVASF, Campus Sede, Av. José de Sá Maniçoba, s/n, Centro, CEP 56304-917, Petrolina, PE, Brasil

*e-mail: jane.gabriel@univasf.edu.br

Received: December 3, 2017 – Accepted: July 17, 2018 – Distributed: February 28, 2020
(With 2 figures)

Abstract

The current study aimed to assess whether the A122V causal polymorphism promotes alterations in the functional and structural proprieties of the CXC chemokine receptor type 1 protein (CXCR1) of cattle *Bos taurus* by *in silico* analyses. Two amino acid sequences of bovine CXCR1 was selected from database UniProtKB/Swiss-Prot: a) non-polymorphic sequence (A7KWG0) with alanine (A) at position 122, and b) polymorphic sequence harboring the A122V polymorphism, substituting alanine by valine (V) at same position. CXCR1 sequences were submitted as input to different Bioinformatics' tools to examine the effects of this polymorphism on functional and structural stabilities, to predict eventual alterations in the 3-D structural modeling, and to estimate the quality and accuracy of the predictive models. The A122V polymorphism exerted tolerable and non-deleterious effects on the polymorphic CXCR1, and the predictive structural model for polymorphic CXCR1 revealed an alpha helix spatial structure typical of a receptor transmembrane polypeptide. Although higher variations in the distances between pairs of amino acid residues at target-positions are detected in the polymorphic CXCR1 protein, more than 97% of the amino acid residues in both models were located in favored and allowed conformational regions in Ramachandran plots. Evidences has supported that the A122V polymorphism in the CXCR1 protein is associated with increased clinical mastitis incidence in dairy cows. Thus, the findings described herein prove that the replacement of the alanine by valine amino acids provokes local conformational changes in the A122V-harboring CXCR1 protein, which could directly affect its post-translational folding mechanisms and biological functionality.

Keywords: CXCR1 protein, *in silico* analysis, protein conservation, single nucleotide polymorphism, structural and functional proprieties.

Predição *in silico* das consequências funcionais e estruturais do polimorfismo de nucleotídeo único não-sinônimo A122V no receptor de quimiocina CXC do tipo I bovino

Resumo

O presente estudo objetivou avaliar se o polimorfismo causal A122V promove alterações nas propriedades funcionais e estruturais da proteína receptora de quimiocina CXC do tipo 1 (CXCR1) de bovino *Bos taurus* por análises *in silico*. Duas sequências de aminoácidos da CXCR1 bovina foram selecionadas a partir do banco de dados UniProtKB/Swiss-Prot: a) sequência não-polimórfica (A7KWG0) contendo alanina (A) na posição 122, e b) sequência polimórfica carregando o polimorfismo A122V, causando a substituição de alanina por valina (V) na mesma posição. As sequências CXCR1 foram analisadas por diferentes ferramentas de Bioinformática para examinar o efeito desse polimorfismo sobre sua estabilidade, função e estrutura, predizer eventuais alterações na sua modelagem estrutural 3-D, bem como estimar a qualidade dos modelos preditos. O polimorfismo A122V exerceu efeitos toleráveis e não-deletérios sobre a CXCR1 polimórfica, apresentando um modelo estrutural de alfa-hélice típico de uma proteína receptora transmembranar para ambas as proteínas. Embora maiores variações nas distâncias entre os pares de aminoácidos nas posições-alvo tenham sido detectadas na proteína polimórfica, mais do que 97% dos aminoácidos em ambos os modelos foram situados em regiões ditas favoráveis e permitidas nos diagramas de Ramachandran. Evidências sustentam que o polimorfismo de nucleotídeo único A122V na proteína receptora CXCR1 está associado à aumentada incidência de mastite clínica em

vacas leiteiras. Assim, as descobertas descritas aqui comprovam que a substituição do aminoácido alanina por valina provoca mudanças conformacionais locais na proteína CXCR1 polimórfica, que podem estar diretamente afetando seus mecanismos de enovelamento pós-traducionais e sua função biológica.

Palavras-chave: proteína CXCR1, análises *in silico*, conservação proteica, polimorfismo de nucleotídeo único, propriedades estruturais e funcionais.

1. Introduction

The CXC chemokine receptor type 1 (CXCR1) characterizes as a class-A, rhodopsin-like G-protein-coupled receptor, the largest class of integral membrane proteins responsible for cellular signal transduction and targeted as drug receptors. Moreover, CXCR1 is one of two high-affinity receptors for the CXC chemokine interleukin-8 (IL-8), a major mediator of immune and inflammatory responses implicated in many disorders, including tumor growth (Emadi et al., 2005). The IL-8, released in response to inflammatory stimuli, binds to the extracellular side of CXCR1, which promotes the neutrophil migration to the site of inflammation (Dakal et al., 2017).

Several genetic polymorphisms in the CXCR1 gene have suggested that this gene is a potential marker of susceptibility to mastitis in dairy cows. Within this context, Zhou et al. (2013) detected four single nucleotide polymorphisms (SNPs) in the CXCR1 gene associated with milk traits in Chinese native cattle. Particularly, the SNP c.365C>T located in exon II of the CXCR1 gene resulted in a non-synonymous mutation [GCC (Ala) > GTC (Val)], which suggests its possible negative effect on the host response against mastitis (Zhou et al., 2013). According to Pokorska et al. (2016), a non-synonymous mutation c.+365T>C in the bovine CXCR1 gene provoked a change in the coded protein [GCC (Ala) to GTC (Val) at the 122nd amino acid], demonstrating stronger association with susceptibility of Polish Holstein-Friesian cows to clinical mastitis.

Over the past decades, powerful experimental approaches of computational biology have been largely raised from development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral and social systems (Ouzounis, 2012). Computational tools are routinely employed to test virtually hypotheses for characterization of genes, determining structural and physicochemical properties of proteins, phylogenetic analyses, and performing simulations to study how biomolecule interacts in a living cell (Mehmood et al., 2014). *In silico* studies of protein and nucleic acid is becoming a useful technique in the field of Bioinformatics, since the knowledge of the 3-D structure of a protein would be an invaluable aid to understand the details of a particular protein (Adzhubei et al., 2010). Within this context, the current study aimed to assess whether the A122V causal polymorphism promotes alterations in the functional and structural proprieties of the CXCR1 protein of bovine *Bos taurus* using computational biology tools. Such *in silico* approach must proportionate new insights

into the protein structure by validating the consequences of the A122V polymorphism in structural and functional stability of the bovine polymorphic CXCR1 protein.

2. Material and Methods

2.1. Search and selection of the sequences of the bovine CXCR1 protein from biological database set

In current study, the sequence of amino acid residues concerning CXCR1 protein of cattle *Bos taurus* (A7KWG0) was searched and selected in FASTA format from non-redundant protein sequence database UniProtKB/Swiss-Prot. This sequence designed non-polymorphic CXCR1 protein has the alanine (A) amino acid at position 122. Moreover, other CXCR1 sequence harboring the A122V causal polymorphism designed polymorphic CXCR1 protein presents the replacement of alanine by valine (V) amino acid at same position.

2.2. Prediction of the effects of the A122V polymorphism on structure, stability and function of the CXCR1 protein

Assessment of the functional impact of the A122V polymorphism on the CXCR1 protein was carried out using Sorting Intolerant From Tolerant (SIFT), which predicts whether an amino acid substitution affects protein function from sequence homology and the physical properties of amino acids (Ng and Henikoff, 2003). Since each substitution in protein-coding regions exhibits potential to affect protein function, SIFT tool involves multistep procedure that takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. Substitutions at each position with normalized probabilities less than a tolerance index of 0.05 are predicted to be intolerant or deleterious; those greater than or equal to 0.05 are predicted to be tolerated (Ng and Henikoff, 2003).

The influence of the A122V polymorphism on the structural and functional stabilities of proteins was predicted using specific empirical rules of the PolyPhen program version 2.0 (Polymorphism Phenotyping v2) that validates possible impacts of amino acid substitutions on the structure and function of human proteins using straightforward physical and evolutionary comparative considerations (Adzhubei et al., 2010). Scores generated by PolyPhen analyses can be classified as “probably damaging” (2.00), “possibly damaging” (1.50-1.99), “potentially damaging” (1.25-1.49), or “benign” (0.00-0.99) (Khan and Ansari, 2017).

Alterations in the polymorphic A122-harboring CXCR1 protein stability was predicted by support vector machine based tool server I-Mutant2.0 that calculates the Gibbs free

energy change (DDG) by subtracting the mutated protein unfold Gibbs free energy from wild type proteins unfold Gibbs free energy (Capriotti et al., 2005). The value of DDG below 0 (<0) indicates decrease in the protein stability and value higher than 0 (>0) indicates increase in protein stability. I-Mutant version 2 also predicts the sign of decrease or increase in Gibbs free energy estimated by Reliability Index (RI) with values ranging of 0 (low estreliability) to 10 (high estreliability).

2.3. Prediction of the 3-D structural models of the CXCR1 proteins by comparative homology modelling analyses

Comparative protein modellings were performed to infer the 3-D structure of the non-polymorphic and polymorphic CXCR1 proteins using Comparative Homology Modelling MODELLER version 9.17 (Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints), providing alignments of the sequences of interest with known related structures (Webb and Sali, 2014). First, the Basic Local Alignment Search Tool BLAST program identified regions of similarity between the CXCR1 query sequences and protein template sequences deposited in the database. Two target-templates were automatically selected: ID 2LNL structure of human CXCR1 in phospholipid bilayers with identity of 79%, total score value of 483 (85%) and e-value of $8e^{173}$, and ID 5LWE crystal structure of the human CC chemokine receptor type 9 in complex with verciron, showing identity of 37%, total score value of 204 (78%) and e-value of $7e^{63}$. According to Vitkup et al. (2001), template models with at least 30% sequence identity are considered 'reasonably accurate' or, for brevity, 'accurate'. To build the 3-D structural models of the proteins of interest by MODELLER analyses, Blast files of the ID 2LNL and ID 5LWE templates were converted to PDB format using Protein Structure Databank available in the Protein Data Bank PDB, a single worldwide repository of 3-D structures of large biological molecules. The building of the 3-D structural models from non-polymorphic and polymorphic CXCR1 sequences was carried out using BLOSUM62 similarity matrix of the MODELLER program, as presented in details by tutorial notes (Eswar et al., 2006). Two evaluative scores are potentially relevant to assess the quality of the predictive 3-D models: GA341, a score derived from statistical potentials with values >0.7 generally indicating a reliable and correctly folded model under high probability (>95%); and zDOPE, normalized Discrete Optimized Protein Energy, an atomic distance-dependent statistical score where negative values indicate better models (Eswar et al., 2006). Finally, homology-modeling analyses of the amino acid side chain conformations was carried out by UCSF Chimera, an extensible molecular modeling system for interactive visualization of high-quality images and animations, and analysis of molecular structures, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles (Pettersen et al., 2004).

2.4. Estimation of the quality of the predictive structural models from non-polymorphic and polymorphic CXCR1 protein sequences

The quality and accuracy of the predictive 3-D models for non-polymorphic and polymorphic CXCR1 proteins was confirmed by Qualitative Model Energy ANalysis QMEAN server, accessing the scoring function QMEANBranE, which calculates specifically trained potentials for three different segments (membrane, interface and soluble) in a transmembrane protein model (Studer et al., 2014). The quality of the predictive models was estimated by QMEAN4 and QMEAN6 Z-scores and for normalization, QMEAN scores of a model are compared to distributions obtained from high-resolution structures solved by X-ray crystallography (Benkert et al., 2011). According to Benkert et al. (2011), higher QMEAN Z-scores indicate better agreement with predictive features and lower mean force potential energy. Other statistical potentials, such as all-atom and C β interactions, solvation and torsion, were also generated by QMEANbrane analyses to assess the overall structural quality of the predictive transmembrane models for non-polymorphic and polymorphic CXCR1.

Furthermore, RAMPAGE server was also employed to validate the structural accuracy and quality of the homology models of the proteins of interest by visualizing energetically allowed regions for backbone dihedral angles of their amino acid residues in Ramachandran plots (Laskowski et al., 2001). In fact, the major effect in the Ramachandran plots corresponds to the presence or absence of the methylene group at C β . In practice, the glycine amino acid has only a hydrogen atom for its side chain with a much smaller van der Waals radius than the CH₃, CH₂, or CH group that starts the side chain of all other amino acids (Ting et al., 2010). On the other hand, the Ramachandran plot for proline, with its 5-membered-ring side chain connecting C α to backbone N, demonstrates a limited number of possible combinations of ψ and ϕ torsion angles (Anderson et al., 2005). A 3-D structural model with $\geq 90\%$ of its residues in the most favored regions of Ramachandran plot is considered as accurate as a 2Å-resolution crystal structure (Laskowski et al., 2001).

3. Results

3.1. Effects of the A122V polymorphism on functional and structural stabilities of the CXCR1 protein

The impact of the A122V substitution on function of the CXCR1 protein by SIFT analyses predicted that this polymorphism is 'tolerable'. Corroborating this finding, prediction of the effect of this polymorphism on the structure and function stability of the protein CXCR1 by PolyPhen-2 analyses demonstrated that p.A122V is 'benign' with score values of 0.001 and 0.000 at levels of variation (sensitivity 0.99 and specificity 0.09) and divergence (sensitivity 1.00 and specificity 0.00), respectively. The polymorphic A122V-harboring CXCR1 sequence was also submitted as input to I-Mutant 2.0, showing DDG and RI values of 0.28 and 3.0, respectively, which indicate increased stability and non-deleterious effect of this nsSNP on structure and function of the polymorphic CXCR1 protein.

3.2. Assessment of the structural variations of the predictive 3-D models from non-polymorphic and polymorphic CXCR1

The predictive 3-D structural model for CXCR1 A122V-harboring protein revealed tridimensional spatial structure typical of a single transmembrane alpha helix of a receptor polypeptide (Figure 1A), exactly similar to predictive structural model for non-polymorphic CXCR1 protein (data not shown). The modeling conformation for both two proteins yielded a characteristic wave-like pattern, reflecting helical structure with breaks in the waves corresponding to helix termini or kinks. Specially, the replacement of alanine by valine amino acid residues revealed local quality score of 0.78 (Figure 1B) in comparison to 0.81 value for non-polymorphic CXCR1 protein (data not shown) at a highly conserved position reached in hydrophobic residues. The GA341 score values for non-polymorphic and polymorphic CXCR1 protein models were calculated in 0.99897 and 0.99611, respectively, whereas DOPE score values were calculated in -35260.26953 and -35282.07813 for non-polymorphic and polymorphic CXCR1 protein models, respectively. Such evaluative parameters indicate a high significance and quality degree of the predictive 3-D models for both proteins in MODELLER results.

Stereochemical alterations in the distances between residues of Val121 and Ala122 amino acids were estimated in 7.573Å for non-polymorphic CXCR1 (Figure 2A), contrarily the distance between Val121 and Val122 amino acids was 8.037Å for polymorphic CXCR1 (Figure 2B). Accordingly, the distance between Ala122 and Ser123 residues was 6.210Å for non-polymorphic CXCR1 (Figure 2C), whereas this distance between residues Val122 and Ser123 was 7.508Å for polymorphic CXCR1 (Figure 2D). Total difference in the distance of the amino acid residues situated at positions 121 and 122 was 0.464Å between non-polymorphic and polymorphic CXCR1 proteins (5.77%), whereas difference in the distance of the amino acids located at positions 122 and 123 was 1.298Å between both sequences (17.29%).

3.3. Estimates of the structural quality and accuracy of the predictive 3-D models for CXCR1 proteins

QMEAN4 and QMEAN6 Z-scores concerning structural quality of the predictive models for A122V-harboring CXCR1 protein was -6.75 and -5.44, respectively, whereas the predictive model for non-polymorphic CXCR1 sequence demonstrated QMEAN4 and QMEAN6 Z-score values of -5.91 and -4.97, respectively. All-atom interactions among chemically distinguishable heavy atoms in local secondary

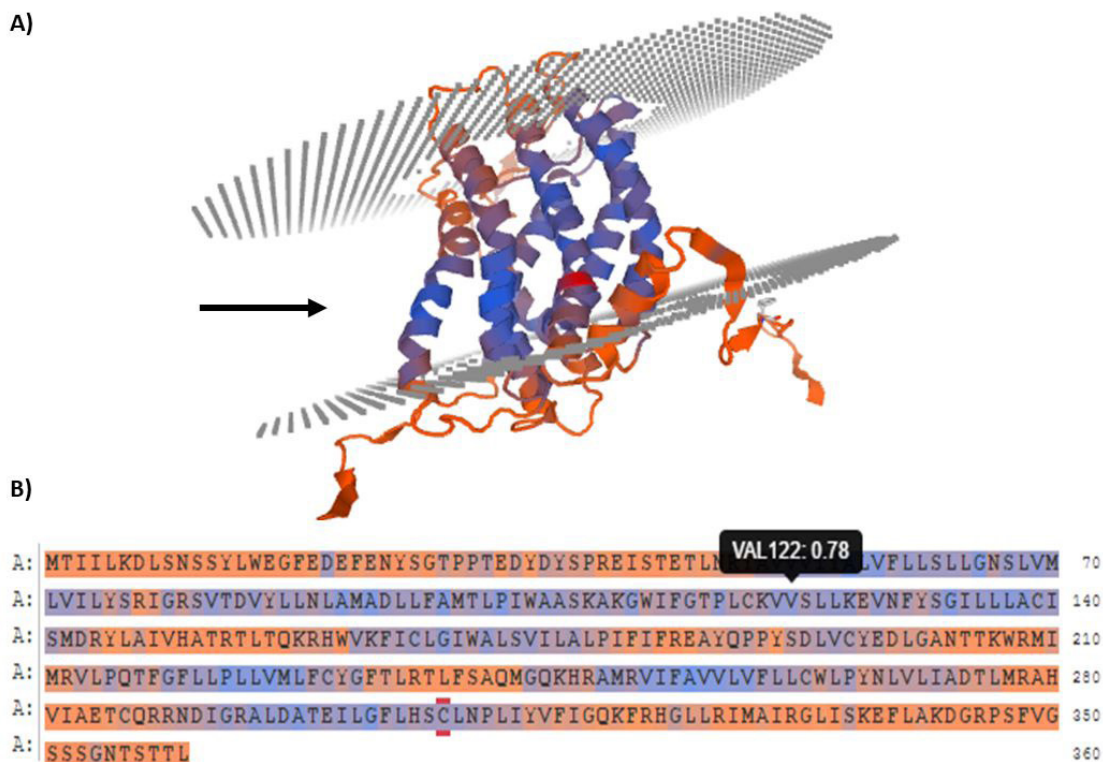


Figure 1. Structural modeling pattern of the polymorphic A122V-harboring CXCR1 protein by QMEANbrane analyses. (A) Tridimensional model typical of a transmembrane receptor protein; (B) Protein sequence with amino acid residues distributed in linear chain. Low and high local quality scores are presented in red (zero-value) and blue (one-value) colors, respectively.

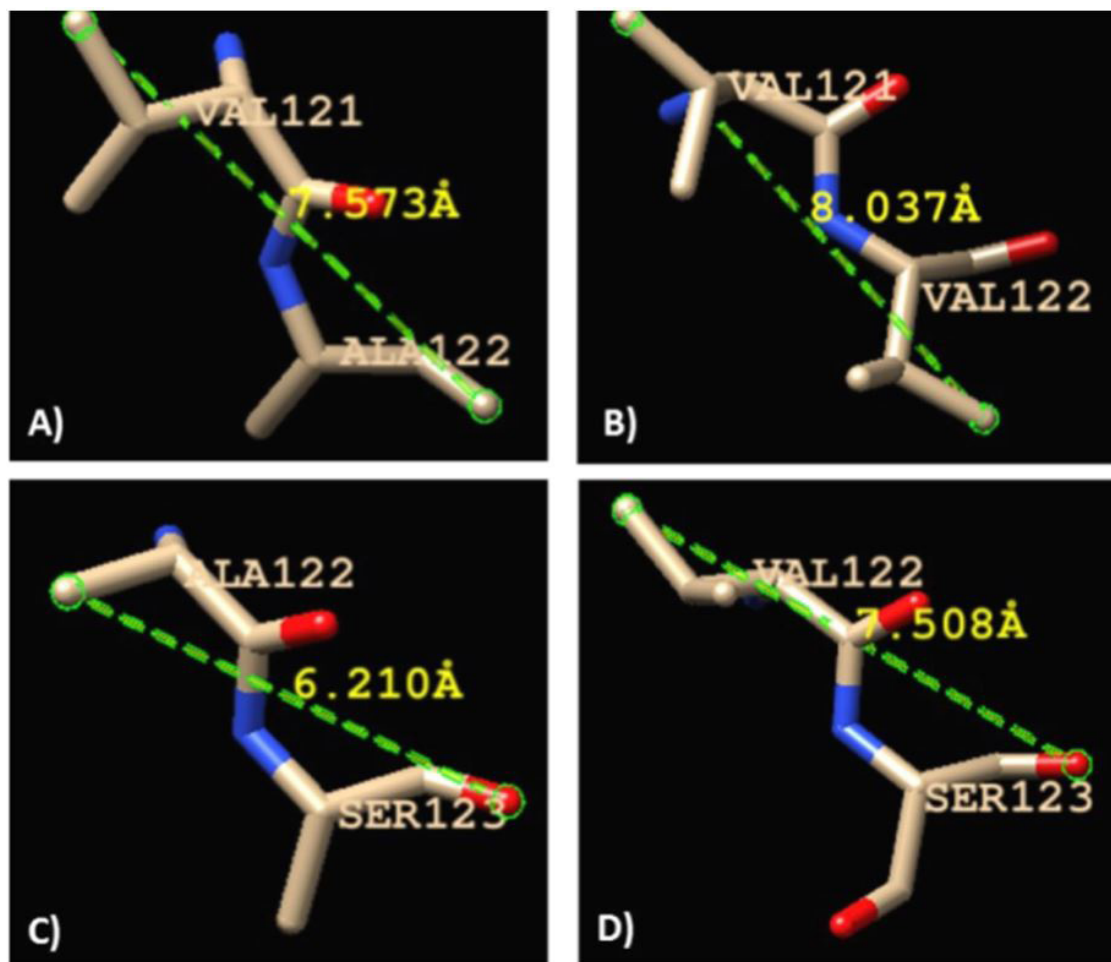


Figure 2. Distances in Å among pairs of amino acid residues situated at target-positions 121, 122 and 123 in the non-polymorphic (A and C) and polymorphic (B and D) CXCR1 proteins by Modeller/Chimera analyses. Amino acid residues: alanine Ala, serine Ser, valine Val.

structure were -6.25 and -6.93 for non-polymorphic and polymorphic CXCR1, respectively, as well as values of interactions among C β positions of the 20 standard amino acids, in particular glycine, were -5.65 and -8.18 in the predictive models for non-polymorphic and polymorphic CXCR1 proteins, respectively. Solvation calculated by counting surrounding atoms around all chemically distinguishable heavy atoms not belonging to the assessed residue itself was -1.73 and -1.97 for non-polymorphic and polymorphic CXCR1 models, respectively. Torsion of the ϕ/ψ central angles of three consecutive amino acids was predicted in -4.43 and -4.78 for non-polymorphic and polymorphic CXCR1 models, respectively. Based on these evaluative parameters, the predictive 3-D models from both CXCR1 protein sequences demonstrated lower modeling quality with strongly negative scores than those generated from non-polymorphic CXCR1.

Furthermore, 93.3% (334 amino acids) and 4.75% of residues (17 amino acids) were situated in the most favorable

and allowed regions, respectively, in Ramachandran plots in the predictive model for non-polymorphic CXCR1 protein. Yet, assessment of the predictive model from polymorphic CXCR1 sequence demonstrated that 92.18% of its residues (330 amino acids) were in the most favorable regions and 21 of residues (5.86%) were situated in allowed regions in Ramachandran plots. Seven amino acids (1.96% of its residues) were situated in disallowed regions for both predictive models. These results prove that the structural quality and accuracy of the predictive 3-D structural models from non-polymorphic and polymorphic CXCR1 protein sequences are perfectly acceptable with good structural modeling quality.

4. Discussion

The bovine A122V-harboring CXCR1 protein sequence was submitted as input to Bioinformatics' tools to infer eventual alterations in its structural and functional stability, demonstrating that this polymorphism exerted benign,

acceptable, tolerable and non-deleterious effects on this protein. Most algorithms for prediction of the functional impact of non-synonymous mutations are based on the observation that evolutionary and structural constraints are non-randomly distributed on proteins (Lee et al., 2009). Several studies have concluded that SIFT and PolyPhen are useful to prioritize changes associated with loss of protein function, and their low specificity to predict gain of protein function should be interpreted with caution to support/refute pathogenicity or functionality of a missense variant (Gnad et al., 2013; Miosge et al., 2015). Thus, findings revealing great stability of the polymorphic CXCR1 protein using Bioinformatics' tools are not sufficient to validate the empirical effects of the A122V polymorphism on its structural and functional stability, since previous report in the literature have described that this polymorphism is yet associated with higher susceptibility to clinical mastitis in dairy cows (Pokorska et al., 2016).

The predictive structural model from polymorphic CXCR1 A122V-harboring protein revealed a tridimensional spatial structure typical of transmembrane alpha helix of a receptor polypeptide, yielding a characteristic wave-like pattern of helical structure with breaks in the waves corresponding to helix termini or kinks (Figure 1A). Transmembrane segments usually characterize as a single alpha helix of a thermodynamically stable transmembrane protein reached by hydrophobic amino acid residues in the interior of the bilayer and the interiors of most proteins of known structure. High local quality score were estimated for polymorphic A122V-harboring CXCR1 protein (Figure 1B), suggesting that the substituting one amino acid for another with similar biochemical properties (nonpolar, hydrophobic amino acids) is less likely to proportionate functional protein alterations. In the meantime, both predictive 3-D models for CXCR1 protein sequences revealed strongly negative scores for evaluative parameters regarding overall quality and accuracy of the structural modeling. Especially for transmembrane proteins, such findings are absolutely expected, since *in silico* Bioinformatics' tools still poorly perform when applied to models of transmembrane segment (Studer et al., 2014). Interestingly, more strongly negative QMEAN Z-scores were predicted for the model of the polymorphic A122V-harboring CXCR1 protein, suggesting that the replacement of the small amino acid alanine by the large amino acid valine could sterically hinder packing during protein folding.

Higher variations in the distances between pairs of amino acid residues at target-positions were demonstrated in the predictive 3-D model for A122V-harboring CXCR1 protein than for non-polymorphic sequence (Figure 2). In the meantime, Ramachandran plots predicted that more than 97% of the amino acid residues in two models were located in favored and allowed conformational regions for the general case (all amino acids except glycine, proline, and pre-proline), for glycine, and for proline. Thus, although remarkable differences in the distance between pairs of amino acid residues are inferred resulting from A122V polymorphism, both predictive CXCR1 models safely

demonstrated high local quality and acceptable structural accuracy. Recently, Guzzi et al. (2017) examined the secondary structure pattern of the bovine A122V-harboring CXCR1 protein, revealing the absence of an alpha helix domain between positions 100 and 150, as possible consequence of the replacement of alanine with high helix-forming propensity by valine with limited helix-forming propensity in the polymorphic protein. Notably, diverse functions of the proteins are directly associated with their appropriate structural folding resulting from physical interactions among their amino acid residues. Recently, Jones et al. (2012) concluded that a pair of residues is spatially close residues in a 3-D structure, if their C α atoms are separated by a minimum distance of at least 6Å. According to Cohen et al. (2009), the side chains of arginine, glutamic acid, valine and leucine amino acids are not symmetrical in respect to the interactions of their head groups. From findings described herein it is reasonable to expect that conformational changes provoked by differences in the rotation of the chain between pairs of amino acids should naturally affect the structural modeling pattern of the polymorphic CXCR1 protein and consequently its biological functionality.

Multiple polymorphisms associated with loss-of-function phenotype have affected the expression and function of the human neuropeptide S receptor as consequence of the disruption of the disulfide bridge through the substitution of a cysteine by a phenylalanine at position 197, promoting profound effect on the overall conformation of the protein and its ligand affinity (Anedda et al., 2011). Theoretical studies have hypothesized that the presence of a rare codon in a silent mutation marked by a synonymous polymorphism in a complex mammalian membrane transport protein alters the substrate specificity; affecting the timing of cotranslational folding and resulting in altered function with consequence to clinically important aspects, as postulated by Kimchi-Sarfaty et al. (2007). Recently, Dakal et al. (2017) showed that some SNPs in the human IL-8 protein provoked local alterations in the structure-function relationship, which could possibly lead to prioritization of diseases, such as cancer. All these experimental evidences support that intricate details of 3-D structures are crucial for protein function and local conformational changes, especially connected to binding sites, may exert stronger negative impacts on the protein function and cause diseases. The disruption of hydrophobic interactions, or the introduction of charged residues into buried sites, or mutations that break beta-sheets often influences severely phenotype and raises the susceptibility for disease (Gong and Blundell, 2010). Within this perspective, Pokorska et al. (2016) speculated that the A122V polymorphism in the CXCR1 protein could result in changes in the receptor function, which may be a reason for increased mastitis incidence in cows carrying this polymorphism. In the meantime, the presence of the alanine and valine residues at position 122 in the bovine CXCR1 protein seems to be very far away from regions of the protein involved in its function (binding sites, active

sites), based on understanding the 3-D structure of the human CXCR1 protein (Park et al., 2012). Accordingly, findings described herein prove that the replacement of the alanine by valine amino acids provoked local conformational changes in the polymorphic A122V-harboring CXCR1 protein, which could directly affect its post-translational folding mechanisms and biological functionality.

Evidences have supported that the CXCR1 receptor is an excellent prospective genetic marker for mastitis resistance in cattle that regulates neutrophil recruitment, migration, killing, and survival during inflammation processes (Dakal et al., 2017) and despite its importance, its molecular mechanism is poorly understood due to the limited structural information available. For the deep understanding of the biological roles attributed to transmembrane proteins as key components of signal transduction, cell-cell adhesion and energy and material transport into and out from the cells, the structure determination of these segments is indispensable. However, only a few transmembrane protein structures have been experimentally determined due to technical difficulties. With the advent of large-scale genomic, increased amounts of sequence information on the proteins and whole proteomes of living organisms have been provided, becoming a crucial challenge of Bioinformatics to elucidate how the structural information should be gained from a sequence. Within this perspective, the findings described herein provide insights into consequences of the A122V non-synonymous single nucleotide polymorphism in the structural and functional proprieties of the bovine CXCR1 protein using computational biology tools, supporting previous practical evidences that this polymorphism is associated with the loss-of-function phenotype increasing the susceptibility of dairy cows to clinical mastitis.

Acknowledgements

This research was supported by the Bioinformatics and Computational Biology Group, designated "BIO in BYTES".

References

ADZHUBEI, I.A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V.E., GERASIMOVA, A., BORK, P., KONDRASHOV, A.S. and SUNYAEV, S.R., 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, vol. 7, no. 4, pp. 248-249. <http://dx.doi.org/10.1038/nmeth0410-248>. PMID:20354512.

ANDERSON, R.J., WENG, Z., CAMPBELL, R.K. and JIANG, X., 2005. Main-chain conformational tendencies of amino acids. *Proteins*, vol. 60, no. 4, pp. 679-689. <http://dx.doi.org/10.1002/prot.20530>. PMID:16021632.

ANEDDA, F., ZUCHELLI, M., SCHEPIS, D., HELLQUIST, A., CORRADO, L., D'ALFONSO, S., ACHOUR, A., MCINERNEY, G., BERTORELLO, A., LÖRDAL, M., BEFRITS, R., BJÖRK, J., BRESSO, F., TÖRKVIST, L., HALFVARSON, J., KERE, J. and D'AMATO, M., 2011. Multiple polymorphisms affect expression and function of the Neuropeptide S Receptor (NPSR1). *PLoS One*, vol. 6, no. 12, pp. e29523. <http://dx.doi.org/10.1371/journal.pone.0029523>. PMID:22216302.

BENKERT, P., BIASINI, M. and SCHWEDE, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, vol. 27, no. 3, pp. 343-350. <http://dx.doi.org/10.1093/bioinformatics/btq662>. PMID:21134891.

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*, vol. 33, pp. W306-10. <http://dx.doi.org/10.1093/nar/gki375>. PMID:15980478.

COHEN, M., POTAPOV, V. and SCHREIBER, G., 2009. Four distances between pairs of amino acids provide a precise description of their interaction. *PLoS Computational Biology*, vol. 5, no. 8, pp. e1000470. <http://dx.doi.org/10.1371/journal.pcbi.1000470>. PMID:19680437.

DAKAL, T.C., KALA, D., DHIMAN, G., YADAV, V., KROKHOTIN, A. and DOKHOLYAN, N.V., 2017. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Scientific Reports*, vol. 7, no. 1, pp. 6525. <http://dx.doi.org/10.1038/s41598-017-06575-4>. PMID:28747718.

EMADI, S., CLAY, D., DESTERKE, C., GUERTON, B., MAQUARRE, E., CHARPENTIER, A., JASMIN, C. and LE BOUSSE-KERDILÈS, M., 2005. IL-8 and its CXCR1 and CXCR2 receptors participate in the control of megakaryocytic proliferation, differentiation, and ploidy in myeloid metaplasia with myelofibrosis. *Blood*, vol. 105, no. 2, pp. 464-473. <http://dx.doi.org/10.1182/blood-2003-12-4415>. PMID:15454487.

ESWAR, N., WEBB, B., MARTI-RENOM, M.A., MADHUSUDHAN, M.S., ERAMIAN, D., SHEN, M.Y., PIEPER, U. and SALI, A., 2006. Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, vol. 15, no. 1, pp. 5.6.1-5.6.30. <http://dx.doi.org/10.1002/0471250953.bi0506s15>. PMID:18428767.

GNAD, F., BAUCOM, A., MUKHYALA, K., MANNING, G. and ZHANG, Z., 2013. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics*, vol. 14, suppl. 3, pp. S7. PMID:23819521.

GONG, S. and BLUNDELL, T.L., 2010. Structural and functional restraints on the occurrence of single amino acid variations in human proteins. *PLoS One*, vol. 5, no. 2, pp. e9186. <http://dx.doi.org/10.1371/journal.pone.0009186>. PMID:20169194.

GUZZI, A.F., OLIVEIRA, F.S.L., AMARO, M.M.S., TAVARES-FILHO, P.F. and GABRIEL, J.E., 2017. Predicting the effects of the single nucleotide polymorphism A122V on the CXC chemokine receptor type 1 of bovine *Bos taurus* (Artiodactyla: Bovidae) by *in silico* analyses. *Biotemas*, vol. 30, no. 4, pp. 1-6. <http://dx.doi.org/10.5007/2175-7925.2017v30n4p1>.

JONES, D.T., BUCHAN, D.W., COZZETTO, D. and PONTIL, M., 2012. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics*, vol. 28, no. 2, pp. 184-190. <http://dx.doi.org/10.1093/bioinformatics/btr638>. PMID:22101153.

KHAN, I. and ANSARI, I.A., 2017. Prediction of a highly deleterious mutation E17K in AKT-1 gene: An *in silico* approach. *Biochemistry and Biophysics Reports*, vol. 10, pp. 260-266. <http://dx.doi.org/10.1016/j.bbrep.2017.04.013>. PMID:29114575.

KIMCHI-SARFATY, C., OH, J.M., KIM, I.W., SAUNA, Z.E., CALCAGNO, A.M., AMBUDKAR, S.V. and GOTTESMAN, M.M., 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, vol. 315, no. 5811, pp. 525-528. <http://dx.doi.org/10.1126/science.1135308>. PMID:17185560.

- LASKOWSKI, R.A., MACARTHUR, M.W. and THORNTON, J.M. 2001. PROCHECK: validation of protein structure coordinates, in International Tables of Crystallography. In: M.G. ROSSMANN and E. ARNOLD, eds. *Crystallography of biological macromolecules*. Dordrecht: Kluwer Academic Publishers, vol. F, pp. 722-725.
- LEE, W., ZHANG, Y., MUKHYALA, K., LAZARUS, R.A. and ZHANG, Z., 2009. Bi-directional SIFT predicts a subset of activating mutations. *PLoS One*, vol. 4, no. 12, pp. e8311. <http://dx.doi.org/10.1371/journal.pone.0008311>. PMID:20011534.
- MEHMOOD, M.A., SEHAR, U. and AHMAD, N., 2014. Use of bioinformatics tools in different spheres of life sciences. *Journal of Data Mining in Genomics & Proteomics*, vol. 5, pp. 158.
- MIOSGE, L.A., FIELD, M.A., SONTANI, Y., CHO, V., JOHNSON, S., PALKOVA, A., BALAKISHNAN, B., LIANG, R., ZHANG, Y., LYON, S., BEUTLER, B., WHITTLE, B., BERTRAM, B.M., ENDERS, A., GOODNOW, C.C. and ANDREWS, T.D., 2015. Comparison of predicted and actual consequences of missense mutations. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 37, pp. E5189-E5198. <http://dx.doi.org/10.1073/pnas.1511585112>. PMID:26269570.
- NG, P.C. and HENIKOFF, S., 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812-3814. <http://dx.doi.org/10.1093/nar/gkg509>. PMID:12824425.
- OUZOUNIS, C.A., 2012. Rise and demise of Bioinformatics? Promise and progress. *PLoS Computational Biology*, vol. 8, no. 4, pp. e1002487. <http://dx.doi.org/10.1371/journal.pcbi.1002487>. PMID:22570600.
- PARK, S.H., DAS, B.B., CASAGRANDE, F., TIAN, Y., NOTHNAGEL, H.J., CHU, M., KIEFER, H., MAIER, K., DE ANGELIS, A.A., MARASSI, F.M. and OPELLA, S.J., 2012. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature*, vol. 491, no. 7426, pp. 779-783. <http://dx.doi.org/10.1038/nature11580>. PMID:23086146.
- PETTERSEN, E.F., GODDARD, T.D., HUANG, C.C., COUCH, G.S., GREENBLATT, D.M., MENG, E.C. and FERRIN, T.E., 2004. UCSF Chimera: a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605-1612. <http://dx.doi.org/10.1002/jcc.20084>. PMID:15264254.
- POKORSKA, J., DUSZA, M., KUŁAJ, D., ŻUKOWSKI, K. and MAKULSKA, J., 2016. Single nucleotide polymorphisms in the CXCR1 gene and its association with clinical mastitis incidence in Polish Holstein-Friesian cows. *Genetics and Molecular Research*, vol. 15, no. 2, pp. 1-8. <http://dx.doi.org/10.4238/gmr.15027247>. PMID:27173275.
- STUDER, G., BIASINI, M. and SCHWEDE, T., 2014. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics*, vol. 30, no. 17, pp. i505-i511. <http://dx.doi.org/10.1093/bioinformatics/btu457>. PMID:25161240.
- TING, D., WANG, G., SHAPOVALOV, M., MITRA, R., JORDAN, M.I. and DUNBRACK JUNIOR, R.L., 2010. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Computational Biology*, vol. 6, no. 4, pp. e1000763. <http://dx.doi.org/10.1371/journal.pcbi.1000763>. PMID:20442867.
- VITKUP, D., MELAMUD, E., MOULT, J. and SANDER, C., 2001. Completeness in structural genomics. *Nature Structural Biology*, vol. 8, no. 6, pp. 559-566. <http://dx.doi.org/10.1038/88640>. PMID:11373627.
- WEBB, B. and SALI, A., 2014. Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, vol. 47, no. 1, pp. 5.6.1-5.6.32. <http://dx.doi.org/10.1002/0471250953.bi0506s47>. PMID:25199792.
- ZHOU, L., WANG, H.M., JU, Z.H., ZHANG, Y., HUANG, J.M., QI, C., HOU, M.H., AN, L.G., ZHONG, J.F. and WANG, C.F., 2013. Association of novel single nucleotide polymorphisms of the CXCR1 gene with the milk performance traits of Chinese native cattle. *Genetics and Molecular Research*, vol. 12, no. 3, pp. 2725-2739. <http://dx.doi.org/10.4238/2013.July.30.10>. PMID:23979897.