Computational prediction of nsSNPs effects on protein function and structure, a prioritization approach for further *in vitro* studies applied to bovine GSTP1

A.H. Falomir-Lockhart[a,*], E.E. Villegas-Castagnaso[a], G. Giovambattista[a], A. Rogberg-Muñoz[a,b]

[a] *IGEVET — Instituto de Genética Veterinaria "Ing. Fernando Noel Dulout" (UNLP-CONICET La Plata), Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina*
[b] *Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina*

A R T I C L E   I N F O

A B S T R A C T

The development of high-throughput technologies in the last decade produced an exponential increase in the amount of biological data available. The case of redox biology and apoptosis is not an exception, and nowadays there is a need to integrate information from multiple "omics" studies. Therefore, validation of proposed discoveries is essential. However, the study in biological systems of the effect of the massive amounts of sequence variation data generated with next-generation sequencing (NGS) technologies can be a very difficult and expensive process. In this context, the present study aimed to demonstrate the advantages of a computational methodology to systematically analyze the structural and functional effects of protein variants, in order to prioritize further studies. This approach stands out for its easy implementation, low costs and low time consumed. First, the possible impact of mutations on protein structure and function was tested by a combination of tools based on evolutionary and structural information. Next, homology modeling was performed to predict and compare the 3D protein structures of unresolved amino acid sequences obtained from genomic resequencing. This analysis applied to the bovine GSTP1 allowed to determine that some of amino acid substitutions may generate important changes in protein structure and function. Moreover, the haplotype analysis highlighted three structure variants worthwhile studying through *in vitro* or *in vivo* experiments.

1. Introduction

The development of high-throughput technologies in the last decade produced an exponential increase in the amount of biological data available. The growth of the "omics" fields allowed the detection of variations in molecular components at different levels [1]; thereby, from one sample it might be obtained information from its genome structure and variants, genes and proteins expression patterns, and metabolites concentration and fluxes, among others [2]. The case of redox biology and apoptosis is not an exception, and nowadays there is a need to integrate information from proteomic, genomic, transcriptomic and metabolomic studies [3–5]. Considering this vast spectrum and amount of information, the increase of false positive signals has been proved and discussed [6,7], hence validation of proposed discoveries is essential [8]. For this task, identifying worthwhile studying cases, separating true signals from false ones, is challenging, and, in this sense, *in silico* studies could help identifying and prioritizing results to be further *in vitro* or *in vivo* validated.

In particular, the next-generation sequencing (NGS) technologies generate massive amounts of information on sequence variability. However, the study of the effect of such great volume of sequence variation data in a biological system can be a very difficult and expensive process. On the contrary, *in silico* analysis methods are easy to perform, cost-effective, reliable and less time-consuming [9], offering a feasible solution to assess the potential impact of amino acid changes. Since these techniques are based on protein structure and sequence, they can predict phenotypic changes arising from genomic variation. Particularly, non-synonymous SNPs (nsSNPs) modify the sequence of proteins and could possibly affect their activity. Therefore, our ability to select nsSNPs for experimental studies can be enhanced by

performing preliminary *in silico* studies, analyzing the potential impact of an amino acid variant on protein structure and function [10].

Within cell redox state, Glutathione S-Transferasa P1 (GSTP1) is a key enzyme. This enzyme is part of the Glutathione S-transferases (GSTs, E.C. 2.5.1.18) superfamily found in organisms from all kingdoms of life, with important biological roles and multiple biomedical and biotechnological applications [11]. One of GSTs major roles is cell protection against xenobiotic substances and products of oxidative stress, conjugating electrophilic and hydrophobic substrates and reactive oxygen species with glutathione (GSH). In addition, GSTs show non-catalytic functions modulating signaling processes that regulate cell proliferation, differentiation and apoptosis [12,13]. Previous studies have shown that GSTP1 haplotypes have differential effects on cellular proliferation and apoptosis [14]. While many studies have focused on correlating the *post mortem* activity of antioxidant enzymes with oxidation processes, the information on the role of GSTP1 on muscle oxidative deterioration and apoptosis is relatively scarce [15,16]. When the animal dies, muscle cells remain alive for a period of time until oxygen and nutrients depletion, herein, the first steps of the *post mortem* conversion of muscle into meat could be used as a model of cellular extreme oxidative stress. In this sense, apoptosis is believed to be the beginning of a complex process that includes a number of physiological and physicochemical changes which occur during and shortly after harvesting [17].

In this research we present a computational methodology to analyze the effects of variants on protein structure and function, to select the haplotype variants to be further studied by *in vitro* biochemical assays. The utility of this method is presented for the case of bovine GSTP1 structural variants in the context of the redox stability and apoptosis related to the conversion of muscle into meat, based on the hypothesized that an alteration in GSTP1 protein activity due to structural variants would lead to differential *post mortem* apoptosis development and hence differential muscle oxidative stability.

2. Materials and methods

2.1. DNA samples, resequencing and haplotype reconstruction

Purified DNA samples were selected from the Veterinary Genetics Institute DNA Sample Bank (IGEVET, UNLP-CONICET; http://www.igevet.gob.ar). DNA was extracted from hair using the NaOH/HCl protocol [18] or blood using the Wizard® Genomic DNA Purification Kit (Promega, USA), following the manufacturer's instructions. The resequencing set comprised 13 bovine DNA samples from several breeds (Angus, Brahman, Creole, Hereford, Holstein, Nellore and Wagyu) to make best use of genetic diversity. The coding regions of the *GSTP1* gene were amplified using the primers and PCR protocols detailed in Supplementary Table 1; they were then sequenced using the sequencing facilities available at IGEVET (http://www.igevet.gob.ar/servicios.html). The obtained sequences were aligned and analyzed to identify the polymorphic positions using DNAMAN v4.15 (Lynnon Corp., Canada). Finally, haplotypes were reconstructed with PHASE v2.1.1 ran with standard parameters [19] and predicted mRNA (p-mRNA) and protein (p-Prot) were obtained for each GSTP1 haplotype.

2.2. Functional analysis of GSTP1 amino acid substitutions

For the functional analysis, several open source tools for sequence- and structure-based approaches were employed to analyze and compare the amino acid substitution effect on the structure, function and stability of the predicted proteins. The reported consensus GSTP1 protein (NP_803482.1) at the NCBI database (https://www.ncbi.nlm.nih.gov/protein) was used as control and for comparison with each of the one amino acid substituted theoretical proteins. The following tools were used:

SIFT - Sorting Intolerant from Tolerant (http://sift.jcvi.org/www/ SIFT_seq_submit2.html) [20] was used to predict the deleterious score of the substituted amino acids according to the conservation of each residue from an evolutionary point of view. Output scores were classified as damaging (0–0.05), potentially damaging (between 0.05 and 0.10), and neutral (above 0.10).

PROVEAN - Protein Variation Effect Analyzer (http://provean.jcvi.org/seq_submit.php) was used to predict the possible deleterious effect of an amino acid substitution on protein structure and function. Protein variants were considered deleterious if the final output score was below the threshold score of −2.5, and neutral if it was above such score.

PolyPhen-2 - Polymorphism Phenotyping v2 (http://genetics.bwh.harvard.edu/pph2/) [21] was used to predict the possible impact of amino acid substitutions on protein structure and function by using straightforward physical and evolutionary comparative considerations. Mutations were classified as probably damaging ($\geq$ 0.85), possibly damaging (0.15–0.85), and benign (0.00– 0.15).

PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp) was used to further analyze the effect of mutations on protein function from an evolutionary point of view. Mutations were classified as probably damaging, when the preservation time of a residue was greater than 450 million years; possibly damaging, when it was between 200 and 450; and probably benign if it was less than 200.

MutPred2 (http://mutpred.mutdb.org/) [22] was used to calculate the probability that an amino acid substitution was pathogenic (over 0.8) or non-pathogenic (below 0.8).

I-Mutant Suite 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) [23] was used to predict protein stability changes due to single-site mutations, according to the change in unfolding Gibbs free energy (DDG) between the wild type and the mutated proteins. Mutations were classified as largely destabilizing protein structure (DDG lower than −0.5 Kcal/mol), largely stabilizing (DDG higher than 0.5 Kcal/mol), or as having a weak effect (DDG between −0.5 and 0.5 Kcal/mol). The input used here was the structure of the reported GSTP1 protein at NCBI, modeled with the I-TASSER server (Iterative Threading ASSEmbly Refinement; https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [24].

HOPE - Have (y)Our Protein Explained (http://www.cmbi.ru.nl/hope/) [25] was used to obtain a prediction and an analysis of the effect of each mutation found on the 3D structure and function of GSTP1. This server performs a BLAST search on several databases such as UniProt and the Protein Data Bank (PDB) to collect structural information and generate the analysis.

2.3. Molecular modeling and visualization

The tertiary structure of the GSTP1 protein was modeled to compare the structures of the native and mutant proteins. Since *Bos taurus* GSTP1 protein tertiary structure is not yet available in the PDB, the I-TASSER server was used to model the 3D structure of the primary sequence of GSTP1 obtained from the NCBI database. Then, haplotypes obtained after the resequencing step were also 3D-modeled with I-TASSER. Afterwards, Swiss PdbViewer (v4.1.0) (http://www.expasy.org/spdbv/) [26] was used for model visualization and image generation. In addition, this software was used to estimate p-Prot volume and surface and the cavity containing the active site. The deviation between haplotypes and the consensus modeled structures was evaluated by their Root Mean Square Deviation (RMSD) values.

3. Results

3.1. Sequencing and haplotype reconstruction summary

Sequence alignment allowed to detect a total of 36 polymorphic positions located in exons (15), 5´UTR region (8) and promoter (13) (Supplementary Figure 1). A total of 24 possible haplotypes were obtained with PHASE, taking into account all the detected SNPs. For

Table 1

Amino acid substitutions predicted with DNA resequencing analysis.

p-Prot	N° of p-mRNA	3	15	33	34	43	44	48	60	65	68	70
Consensus sequence[a]		P	C	V	V	P	L	C	D	Q	A	L
I	4	–	–	–	–	–	–	–	–	–	–	–
II	1	–	–	–	–	–	–	–	–	–	–	Q
III	1	–	–	–	L	Q	F	G	–	–	–	–
IV	1	–	–	F	–	–	–	–	–	–	–	–
V	1	–	W	–	–	–	–	–	–	–	–	–
VI	2	–	–	–	–	–	–	–	–	–	T	–
VII	1	S	–	–	–	–	F	–	N	H	–	Q

p-Prot: predicted protein; **p-mRNA**: predicted mRNA.

　[a] Consensus protein sequence from the NCBI database NP_803482.1.

protein diversity analysis, only SNPs located in the exonic regions were considered, obtaining a total of 11 different p-mRNA (Supplementary Table 2). Further, the analysis of the corresponding amino acid sequence for those p-mRNAs resulted in seven possible translated primary structures (p-Prot) for GSTP1 protein, including 11 non-synonymous amino acid substitutions distributed in the different breeds when compared to the reported sequence of this protein (Table 1).

3.2. Effect of the predicted non-synonymous amino acid changes on structure and stability

To determine whether the identified non-synonymous amino acid substitutions affected GSTP1 function or stability, we further analyzed these variants using six *in silico* prediction algorithms: SIFT, PROVEAN, PolyPhen-2, PANTHER, MutPred 2 and I-Mutant 3.0. Table 2 summarizes the results obtained for each amino acid substitution analyzed individually. We also used Project HOPE to have an insight of each amino acid substitution and of the possible effect on protein structure and function. It is important to remark that these are all analysis of single-site mutations comparing the stability of a theoretically one mutated amino acid protein with the consensus.

To improve the strength of our analysis, data were evaluated by combining all the tools used. Briefly, SIFT and PROVEAN categorized the variants equivalently. The six variants detected by SIFT as damaging or potentially damaging (C15W, C48G, Q65H, L70Q, V33F and D60N) were also detected as deleterious by PROVEAN; the rest were neutral with both tools. PolyPhen-2 and PANTHER also detected four and three variants, respectively, of those six previously detected. PolyPhen-2 detected a new probably damaging substitution (A68T), while PANTHER classified two new substitutions as possibly damaging (V33F and L44F). In addition, we performed I-Mutant 3.0 analysis of GSTP1 in order to add another layer of refinement to the amino acid

substitution characterization, by considering the DDG change between the native and mutated protein. This analysis yielded a significant decrease of almost all of the substitutions and the highest change (− 2.58 Kcal/mol) for L70Q.

In each algorithm, distinct parameters were used to evaluate the amino acid substitutions; the higher the number of significant results for the substitution, the more likely they genuinely affected the structure and function of the protein. Thereby, whereas some substitutions did not apparently affect (P3S, P43Q, L44F) or had a low influence (V33F, V34L) on GSTP1 structure and stability, others exerted a stronger influence (C15W, C48G, D60N, Q65H, A68T, L70Q).

3.3. Assessment of the effects of mutations on tertiary protein structure and function

In order to further the study of these amino acid substitutions and analyze their individual impact on protein structure and function, we performed subsequent HOPE analysis. The results showed that in most comparisons, the wild-type and mutant residues differed in their properties.

All the possible substitutions occurred in the GSTP1 N-terminal domain, where the binding site for GSH is located; from those substitutions, eight were also in contact with another domain of the protein (P3S, V33F, V34L, P43Q, L44F, C48G, D60N, L70Q). Hence their substitution may modify the interactions between those two domains. To summarize, most of the amino acids were substituted with larger residues (with the exception of P3S and C48G), leading to possible steric problems and differences in protein folding. Furthermore, some residues were substituted with less hydrophobic amino acids (P3S, P43Q, C48G, A68T, L70Q), which would generate a possible loss of hydrophobic interactions. On the other hand, since some residues were located near (C15W and L44F) or directly involved (Q65H) in GSH binding, their substitution could disturb the local structure and therefore protein activity. Finally, primary structure analysis of homologous proteins showed that many of these positions were not conserved, some of the new residues had similar properties to the homologous protein alternative residues (V34L, P43Q, L70Q), and some had different properties (P3S, V33F, C48G). In two cases, the mutant residue appeared as a possible alternative amino acid (L44F and D60N), making a substitution more likely to occur.

3.4. Protein structure homology modeling and analysis

So far, all substitutions were analyzed individually. Therefore, in order to compare the accumulated effect of the different mutations, the observed DNA sequencing predicted haplotypes were considered. As

Table 2

In silico analysis of the individual alternative amino acid substitutions detected in the resequencing stage, using different web-based tools.

Aminoacid position		3	15	33	34	43	44	48	60	65	68	70
NCBI Reference NP_803482.1		P	C	V	V	P	L	C	D	Q	A	L
Alternative amino acid		S	W	F	L	Q	F	G	N	H	T	Q
Web-based tools	SIFT[a]	0.3	0.02*	0.1*	0.12	0.5	0.23	0.05*	0.06	0*	0.34	0*
	PROVEAN[b]	− 2.135	− 5.652*	− 3.459*	− 2.043	0.162	− 1.498	− 7.753*	− 3.634*	− 4.704*	− 2.102	− 5.315*
	PolyPhen-2[c]	0.006	1*	0.097	0.001	0.018	0.017	0.279*	0.021	1*	0.979*	1*
	PANTHER[d] (millions of years)	176	176	361	750*	86	361	750*	750*	750*	1629*	750*
	MutPred 2[e]	0.483	0.898*	0.279	0.220	0.236	0.121	0.822*	0.441	0.897*	0.829*	0.928*
	I-Mutant 3.0[f](Kcal/mol)	− 1.18*	− 1.02*	− 1.01*	− 0.81*	− 0.95*	− 0.88*	− 1.21*	− 0.06	− 1.15*	− 0.75*	− 2.58*

　* Highlights the significant predictions from each web-based tool.

　[a] p-value < 0.05: damaging; 0.05–0.10: potentially damaging; > 0.10: neutral.

　[b] score < − 2.5: deleterious; > − 2.5: neutral.

　[c] probability > 0.85: probably damaging; 0.15–0.85: possibly damaging; < 0.15: benign.

　[d] preservation time > 450: probably damaging; 200–450: possibly damaging; < 200; probably benign.

　[e] probability > 0.8: pathogenic; < 0.8: non-pathogenic.

　[f] Unfolding Gibbs free energy change (DDG) < -0.5: large decrease of stability; − 0.5–0.5: neutral; > 0.5: large increase of stability.

Table 3

Surface and volume of the modeled proteins and active site, and Root mean square deviation (RMSD) of the models against the consensus.

p-Prot	Protein		Active site		Backbone RMSD
	Surface	Volume	Surface	Volume	
I	9493	25,525	743	741	–
II	9142	25,782	818	762	0.378
III	9365	25,802	884	1102	0.473
IV	9331	25,673	745	757	0.411
V	9211	25,769	701	650	0.448
VI	9467	25,751	469	464	0.486
VII	8975	26,004	666	633	0.419
5GSS_A	8969	25,119	858	794	0.595

Backbone RMSD is the average difference in the position of the protein backbone atoms compared with the consensus. Consensus (p-Prot I).

previously described, seven different translated proteins were obtained from the 11 different haplotypes due to synonymous mutations.

Unfortunately, no crystallized structure of bovine GSTP1 was publically available in the PDB at the moment of this analysis. Therefore, all the 3D structures were modeled using the I-TASSER server to compare the wild-type and mutant proteins. For wild-type GSTP1, the protein sequence available in the NCBI database (NP_803482.1) was used, whereas for the mutant proteins, the seven p-Prot translated from the eleven p-mRNAs were used. Despite I-TASSER is configured to show five alternative models with the best score, only one structure was generated for all seven p-Prot, indicating how conserved the structure of this protein family is. Furthermore, the alignment of the GSTP1 primary sequence from different species showed little residue variations, and the GSH binding amino acids were particularly highly conserved (Supplementary Figure 2).

The structures modeled were further analyzed using the Swiss PdbViewer software. The protein and active site areas and volumes were determined, as well as the RMSD of the backbone atoms of each p-Prot from those of the modeled consensus sequence (Table 3). Since there is no bovine GSTP1 structure published to compare, the human crystallized structure 5GSS_A was used instead. The 5GSS_A structure has an 86% sequence identity with the bovine GSTP1, with an additional 11% if amino acid similarity is considered, meaning it is a very similar protein. Briefly, little variation was found in total protein surface and volume, whilst more notorious changes were detected in surface and volume of the cavity containing the active site.

4. Discussion

As was set there is a need to differentiate the false positive results and prioritize variants for the biochemical analysis [8]. Analysis of the NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp) showed that genes coding these enzymes present a high number of polymorphic positions, many of which occur in coding regions. Accordingly, in this research a total of 128 SNPs were found in the bovine GSTP1 gene (mutation rate, one SNP every 25.6 bp), which was higher than the approximated rate of one SNP every 100 bp reported [27]. From those variants, 36 were located within the coding regions (mutation rate, one SNP every 30.2 bp), demanding a prioritization to adequate and optimize resources when deciding further *in vitro* or *in vivo* studies. Additionally, a high diversity of the DNA sequence was observed when comparing the gene in different species, despite, a high degree of conservation was observed in protein primary sequence (Supplementary Figure 2) suggesting that the genetic variation should not affect greatly the protein to persist.

Within all the genomic information, non-synonymous SNPs (nsSNPs) are important factors contributing to the functional diversity of the encoded proteins within and between populations. In this sense, prediction of the functional consequences of genetic variations would

help understand the molecular mechanisms underlying the observed phenotypic variability. Synonymous SNPs could lead to RNA molecules with possible different translation rates, which would have an impact on protein folding, yielding altered conformations and ultimately varying total enzyme activity [28]. In this study, we analyzed the protein predicted tertiary structure, without considering synonymous mutations since they would not produce differences in the primary structure of proteins. Likewise, SNPs located in the promoter/regulatory region and introns may also have an effect on the phenotype due to differential expression levels and altered mRNA conformation [29,30]. Thereby, performance of wet laboratory experiments to study the biological significance of each SNP would be a highly laborious task as the great number of SNPs and their combinations would lead to an exponential number of expensive and time-consuming experiments. Nowadays, bioinformatics tools can be used for prioritization through a better understanding of the effects caused by those genomic variants and the screening of potentially deleterious SNPs, before conducting further investigations [31].

Firstly the effect of each nsSNP can be analyzed to detect which of them would have an effect on the protein structure and function. For that task, the genetic variants can be studied using sequence- and structure-based algorithms, such as SIFT, PROVEAN, PolyPhen-2, PANTHER, MutPred 2 and I-Mutant 3.0. This analysis provides information from evolutionary and energetic points of view, analyzing the conservation of residues in comparison to multiple alignments. From the functional analysis, concordant results of all the tools indicated that C15W, C48G, Q65H and L70Q may have an impact on protein structure or function. Structure-based prediction using I-Mutant 3.0 also showed that almost all possible amino acid substitutions were highly destabilizing (only D60N resulted in a neutral change in protein stability), in agreement with the high degree of sequence conservation of this protein among different species. Results of *in silico* analysis using HOPE were in agreement as well. As this enzyme acts non-enzymatically in the apoptotic pathways [13] any structural variation could lead to changes in the regulation of these processes. Additionally, since all substitutions were located in the N-terminal domain (Supplementary Figure 3), they would be important for the binding of other molecules and the activity of GSTP1, especially glutamine in Q65H, which is directly involved in GSH binding. It is worth noting that, besides binding GSH, the residues around the pocket containing the active site define the specificity for the second substrate. This suggests that changes in molecular interactions and misfolding can affect protein structure and function, and that these variants might be functionally involved in differential oxidative stability and apoptosis development.

A second step in this methodology is studying the haplotypes, in order to evaluate the combined effect of the variants. In this sense, the modeling of the tertiary structure of proteins allows to further understand the effect of the possible amino acid substitutions. To begin with, the consensus sequence of the protein was observed in all the breeds studied (Table 1). Furthermore, its predicted structure showed an active site volume and surface slightly deviated from the human crystallized protein. Regarding the rest of p-Prot, four presented only one amino acid substitution (II, IV, V and VI), while p-Prot III and VII presented multiple amino acid substitutions. The substitution IV (V33F), where valine was substituted by a bigger but still hydrophobic phenylalanine, it may not result in a relevant change as this residue was protruding outwards on the surface of the structure. While p-Prot II and V structures presented substitutions located inside the modeled structures, these may influence protein folding, damaging protein structure and function. In p-Prot V (C15W), the small hydrophilic cysteine is substituted by a big, hydrophobic tryptophan that may generate steric problems, evidenced in the reduction in surface and volume of the active site. On the other hand for p-Prot II (L70Q) the appearance of a polar glutamine may generate a disruption of the interaction within the core. On the contrary, the analysis of the modeled structure of p-Prot VI

showed that it had the greatest deviation from the consensus according to RMSD results due to a reduction in the active site volume and surface by almost a half. The A68T substitution produced a change in amino acid properties, as the nonpolar alanine was replaced by a threonine capable of generating new hydrogen bonds in the surroundings of the active site (Supplementary Figure 4).

In the case of p-Prot III and VII, they presented multiple amino acid substitutions, which reflected in the structure results. In p-Prot III, the increased volume of the active site may be explained by the substitution of the rigid proline in P43Q and the introduction of the flexible glycine in C48G, in an α-helix near the active site (Supplementary Figure 5a). This structure is located in an important loop that connects the two typical structural motifs of a thioredoxin-like fold that constitutes the structure of GSTs N-terminal domain [32]. Even though not predicted as damaging, the rest of the substitutions in this structure were also located in this loop (Supplementary Figure 5b), possibly contributing to the observed differences in active site surface and volume. On the other hand, p-Prot VII included probably the most important substitution, Q65H, affecting a residue directly involved in GSH binding. Furthermore, the surface and volume of the active site were greatly reduced, making this an interesting variant to be studied *in vitro*, since a change in substrate binding, and therefore in enzymatic activity, would be expected.

In summary, the presented methodology allowed to determine that some of the predicted amino acid substitutions may generate important changes in protein structure and function, and are good candidates to be further studied in *in vitro* or *in vivo* experiments. These changes could affect either the enzymatic activity, by altering the affinity for the substrates, the reaction speed or the release of the products, or the non-enzymatic function, trough a protein-protein interaction variation. Moreover, the haplotype analysis enhanced this computational approach highlighting that the predicted structures III, VI and VII presented the greatest deviations in the protein structure or the active site volume and surface, thus becoming prioritized structures for further studies.

5. Conclusions

The present study aimed to demonstrate the advantages of a computational methodology to systematically analyze the structural and functional effects of protein variants, in order to prioritize further studies. This approach stands out for its easy implementation, low costs and low time consumed. First, the possible impact of mutations on protein structure and function was tested by a combination of tools based on evolutionary and structural information. Next, homology modeling was performed to predict and compare the 3D protein structures of unresolved amino acid sequences obtained from genomic resequencing. This analysis applied to the bovine GSTP1 highlighted three structure variants worthwhile studying *in vitro*.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.freeradbiomed.2018.10.403.

References

[1] M.V. Schneider, S. Orchard, Omics technologies, data and bioinformatics principles, in: B. Mayer (Ed.), Bioinformatics Omi Data, Humana Press, Totowa, NJ, 2011, pp. 3–30, , https://doi.org/10.1007/978-1-61779-027-0_1.

[2] D. Field, S.-A. Sansone, A. Collis, T. Booth, P. Dukes, S.K. Gregurick, K. Kennedy, P. Kolar, E. Kolker, M. Maxon, S. Millard, A.-M. Mugabushaka, N. Perrin, J.E. Remacle, K. Remington, P. Rocca-Serra, C.F. Taylor, M. Thorley, B. Tiwari, J. Wilbanks, Omics Data Sharing, Sci. (80-.) 326 (2009) 234–236, https://doi.org/10.1126/science.1180598.

[3] J.C. Reed, K.S. Doctor, A. Godzik, The domains of apoptosis: a genomics perspective, Sci. Stke. 2004 (2004) re9, https://doi.org/10.1126/stke.2392004re9.

[4] M. Thamsen, U. Jakob, The redoxome: proteomic analysis of cellular redox networks, Curr. Opin. Chem. Biol. 15 (2011) 113–119, https://doi.org/10.1016/J.CBPA.2010.11.013.

[5] R.-S. Wang, W.M. Oldham, B.A. Maron, J. Loscalzo, Systems biology approaches to redox metabolism in stress and disease states, Antioxid. Redox Signal. 29 (2018) 953–972, https://doi.org/10.1089/ars.2017.7256.

[6] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B. 57 (1995) 289–300 ⟨http://www.jstor.org/stable/2346101⟩.

[7] M.J. Marino, How often should we expect to be wrong? Statistical power, P values, and the expected prevalence of false discoveries, Biochem. Pharmacol. 151 (2018) 226–233, https://doi.org/10.1016/J.BCP.2017.12.011.

[8] J.P.A. Ioannidis, M.J. Khoury, Improving validation practices in "omics" research, Science 334 (2011) 1230–1232, https://doi.org/10.1126/science.1211811.

[9] P.K. Singh, K.N. Mistry, A computational approach to determine susceptibility to cancer by evaluating the deleterious effect of nsSNP in XRCC1 gene on binding interaction of XRCC1 protein with ligase III, Gene 576 (2016) 141–149, https://doi.org/10.1016/j.gene.2015.09.084.

[10] T.A. Masoodi, V. Rao Talluri, N.A. Shaik, J.Y. Al-Aama, Q. Hasan, Functional genomics based prioritization of potential nsSNPs in EPHX1, GSTT1, GSTM1 and GSTP1 genes for breast cancer susceptibility studies, Genomics 99 (2012) 330–339, https://doi.org/10.1016/j.ygeno.2012.04.006.

[11] M. Kosloff, G.W. Han, S.S. Krishna, R. Schwarzenbacher, M. Fasnacht, M.-A. Elsliger, P. Abdubek, S. Agarwalla, E. Ambing, T. Astakhova, H.L. Axelrod, J.M. Canaves, D. Carlton, H.-J. Chiu, T. Clayton, M. DiDonato, L. Duan, J. Feuerhelm, C. Grittini, S.K. Grzechnik, J. Hale, E. Hampton, J. Haugen, L. Jaroszewski, K.K. Jin, H. Johnson, H.E. Klock, M.W. Knuth, E. Koesema, A. Kreusch, P. Kuhn, I. Levin, D. McMullan, M.D. Miller, A.T. Morse, K. Moy, E. Nigoghossian, L. Okach, S. Oommachen, R. Page, J. Paulsen, K. Quijano, R. Reyes, C.L. Rife, E. Sims, G. Spraggon, V. Sridhar, R.C. Stevens, H. van den Bedem, J. Velasquez, A. White, G. Wolf, Q. Xu, K.O. Hodgson, J. Wooley, A.M. Deacon, A. Godzik, S.A. Lesley, I.A. Wilson, Comparative structural analysis of a novel glutathione S -transferase (ATU5508) from Agrobacterium tumefaciens at 2.0 Å resolution, Proteins Struct. Funct. Bioinform. 65 (2006) 527–537, https://doi.org/10.1002/prot.21130.

[12] G. Di Pietro, L.A.V. Magno, F. Rios-Santos, Glutathione S-transferases: an overview in cancer research, Expert Opin. Drug Metab. Toxicol. 6 (2010) 153–170, https://doi.org/10.1517/17425250903427980.

[13] E. Laborde, Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death, Cell Death Differ. 17 (2010) 1373–1380, https://doi.org/10.1038/cdd.2010.80.

[14] S.L. Holley, A.A. Fryer, J.W. Haycock, S.E.W. Grubb, R.C. Strange, P.R. Hoban, Differential effects of glutathione S-transferase pi (GSTP1) haplotypes on cell proliferation and apoptosis, Carcinogenesis 28 (2007) 2268–2273, https://doi.org/10.1093/carcin/bgm135.

[15] U. Pastsart, M. De Boever, E. Claeys, S. De Smet, Effect of muscle and post-mortem rate of pH and temperature fall on antioxidant enzyme activities in beef, Meat Sci. 93 (2013) 681–686, https://doi.org/10.1016/j.meatsci.2012.11.008.

[16] S.B. Rønning, P.V. Andersen, M.E. Pedersen, K. Hollung, Primary bovine skeletal muscle cells enters apoptosis rapidly via the intrinsic pathway when available oxygen is removed, PLoS One 12 (2017) e0182928, https://doi.org/10.1371/journal.pone.0182928.

[17] C.M. Kemp, T. Parr, Advances in apoptotic mediated proteolysis in meat tenderisation, Meat Sci. 92 (2012) 252–259, https://doi.org/10.1016/j.meatsci.2012.03.013.

[18] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, ⟨https://www.cabdirect.org/cabdirect/abstract/19901616061⟩ (Accessed 16 January 2018).

[19] M. Stephens, P. Scheet, Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation, Am. J. Hum. Genet. 76 (2005) 449–462 ⟨http://stephenslab.uchicago.edu/assets/papers/Stephens2005.pdf⟩ (accessed 22 January 2018).

[20] P.C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions, Genome Res. 11 (2001) 863–874, https://doi.org/10.1101/gr.176601.

[21] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat. Methods 7 (2010) 248–249, https://doi.org/10.1038/nmeth0410-248.

[22] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H.-J. Nam, M. Mort, D.N. Cooper, J. Sebat, L.M. Iakoucheva, S.D. Mooney, P. Radivojac, MutPred2: inferring the molecular and phenotypic impact of amino acid variants, BioRxiv (2017) 1–28, https://doi.org/10.1101/134981.

[23] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon

mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005) W306–W310, https://doi.org/10.1093/nar/gki375.

[24] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (2015) 7–8, https://doi.org/10. 1038/nmeth.3213.

[25] H. Venselaar, T.A.H. te Beek, R.K.P. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinform. 11 (2010) 548, https://doi. org/10.1186/1471-2105-11-548.

[26] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb viewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723, https:// doi.org/10.1002/elps.1150181505.

[27] 1000 Bull Genomes Project. ⟨www.1000bullgenomes.com⟩ (Accessed 14 April 2018).

[28] C. Kimchi-Sarfaty, J.M. Oh, I.-W. Kim, Z.E. Sauna, A.M. Calcagno, S.V. Ambudkar, M.M. Gottesman, A "silent" polymorphism in the MDR1 gene changes substrate specificity, Sci. (80-.) 315 (2007) 525–528, https://doi.org/10.1126/science. 1135308.

[29] J. Ramírez-Bello, M. Jiménez-Morales, Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multi-factorial diseases, Gac. Med. Mex. 153 (2017) 218–229 ⟨http://www.ncbi.nlm.nih. gov/pubmed/28474710⟩ (accessed 15 February 2018).

[30] B. Shastry, SNPs: impact on gene function and phenotype, Methods Mol. Biol. (2009) 3–22, https://doi.org/10.1007/978-1-60327-411-1_1.

[31] J.T.L. Mah, E.S.H. Low, E. Lee, In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery, Drug Discov. Today 16 (2011) 800–809, https://doi.org/10.1016/j.drudis.2011.07.005.

[32] C. Frova, Glutathione transferases in the genomics era: new insights and perspectives, Biomol. Eng. 23 (2006) 149–169, https://doi.org/10.1016/j.bioeng.2006.05. 020.