sciendo

Nova Biotechnol Chim (2019) 18(2): 115-123
DOI: 10.2478/nbec-2019-0014

Nova Biotechnologica *et* Chimica

Deleterious amino acid substitutions with a series of putative damaging effects on egg components are revealed in the ovalbumin gene family; an *in silico* approach

Mohammed Baqur S. Al-Shuhaib

Department of Animal Production, College of Agriculture, Al-Qasim Green University, 8-Al Qasim, Babil 51001, Iraq

Article info

Article history:
Received: 18th January 2019
Accepted: 14th April 2019

Keywords:
Chickens
Egg production
Oviduct
Prediction
Single Nucleotide Polymorphism

Abstract

This study was conducted to identify the most deleterious nonsynonymous single nucleotide polymorphisms (nsSNPs) in the ovalbumin gene family, including *OVALX*, *OVALY*, and *OVAL* genes, which are involved in the synthesis of the most important components in the chickens' eggs using a comprehensive *in silico* approach. Ten different computational servers were utilized to prioritize the possible deleterious effects of the retrieved nsSNPs in terms of structure, function, and stability. Results indicated entirely damaging effects of *H365P* in *OVALX*, *I167T* in *OVALY*, and *V209G*, *L231P*, *F307C*, and *S317P* in *OVAL* proteins. Further prediction tools showed that all of these deleterious nsSNPs were positioned in variable locations within several α-helix motifs in all studied ovalbumin proteins. Furthermore, all witnessed nsSNPs were predicted to be resided in the receptors binding sites, signifying remarkable involvement of such nsSNPs in damaging of the altered proteins. In conclusion, the present study provides the first inclusive data with regard to the most deleterious nsSNPs in *OVALX*, *OVALY* and *OVAL* genes in chickens. The present bioinformatics data may be useful for breeders who intend to raise chickens for egg production, in such a way the presence of any of these deleterious nsSNPs in any selected breed may possess several damaging effects on the egg components, which may impair egg production. Therefore, it can be stated that breeders have to confirm the absence of any of these deleterious nsSNPs before being proceeded further for large-scale egg-production purposes.

University of SS. Cyril and Methodius in Trnava

Introduction

The ovalbumin gene family in *Gallus gallus* is composed of three homologous genes located on chromosome 2, namely ovalbumin (*OVAL*), ovalbumin-related protein X (*OVALX*), and ovalbumin-related protein Y (*OVALY*). *OVAL*, *OVALX*, and *OVALY* proteins belong to the serine protease inhibitor (serpin) family whose members share the same overall tertiary structure of eight to nine α-helices and three beta-sheets. All these three genes are mainly expressed in the oviduct (Sugimoto *et al.* 2001), and more specifically by tubular gland cells of the chicken's magnum (Rehault-Godbert *et al*. 2013). Sequences alignment of these three genes have shown a sequence identity of 73 % between *OVAL* and *OVALX*, of 72 % between *OVAL* and *OVALY*, and 82 % between *OVALX* and *OVALY* genes (Da Silva *et al.* 2015). Although *OVAL*, *OVALY*, and *OVALX* proteins have exerted high homology in their primary sequences, distinct and subtle physicochemical and structural differences have exhibited among them that might be associated

Corresponding author: baquralhilly_79@yahoo.com; mohammed79@agre.uoqasim.edu.iq

with specific function on each one. In addition to their major localization in the egg white (Hashim *et al*. 2019), *OVAL*, *OVALX*, and *OVALY* proteins have also been detected in all other egg compartments, including eggshell, egg yolk, and vitelline membrane (Mann *et al*. 2008).

OVAL is the major egg white protein and might be a source of amino acids for the developing embryo (Liu *et al*. 2015). *OVAL* is synthesized in the tubular gland cells in the hen's oviduct. It is well known that such a synthesis mechanism is under hormonal control with a consequent effect on egg white formation. *OVAL* accounts for about 54 % of the total proteins of egg white and more than 80 % of the total proteins of egg yolk (Rehault-Godbert *et al*. 2018). However, the current hypothesis is that *OVAL* serves as a source of amino acids for the developing embryo. The synthesis of *OVAL* is upregulated in the magnum in the early onset of the post-ovulation period (Zhao *et al*. 2016). *OVALX* gene and its closest neighbour *OVALY* gene, are highly similar genes and have a close relation with *OVAL*. Both have likely arisen by duplication events from a common ancestral *OVAL* gene because the two gene-coding sequences share highly significant similarity with that of the *OVAL* (Da Silva *et al*. 2015). Though *OVALX* and *OVALY* properties have been described recently, their biological functions remain to be undefined. With regard to *OVALX*, it has been shown to exhibit antimicrobial activities against pathogens via its positively charged heparin-binding domain exposed at its surface (Rehault-Godbert *et al*. 2013). These activities suggest considerable participation for *OVALX* in egg innate defense mechanisms (Dombre *et al*. 2017; DaSilva *et al*. 2019). Meanwhile, the role of *OVALY* has not been explored yet but because its physiological function is not known. However, increasing data in the literature have shown that its abundance was affected during incubation, suggesting a potential involvement of *OVALY* in some aspects of fertilization or embryonic development (Qiu *et al*. 2012).

Thus, any deleterious amino acid substitution in these proteins may have a series of damaging effects on their biological activity leading to a series of negative consequences on eggs formation. Irrespective of ovalbumin genes, the recent computations of amino acid substitutions have been applied on several genetic markers to predict a variety of physiological effects in birds, such as chickens (Yakubu *et al*. 2017), ostriches (Al-Shuhaib *et al*. 2018), and quails (Al-Shuhaib *et al*. 2019). Up to date, no comprehensive analyses have presented to highlight this aspect in *OVAL*, *OVALX*, and *OVALY* genes and their altered proteins. Therefore, this study aims to exploit the advent of computational biology to present the most deleterious nonsynonymous single nucleotide polymorphisms (nsSNPs) that may be involved in preventing of *OVAL*, *OVALX*, and *OVALY* proteins in undergoing their roles in the proper egg white formation.

Experimental

Molecular modelling

The uniprotKB codes for *OVAL*, *OVALX*, and *OVALY* proteins were A0A2H4Y7U9, A0A1D5PI58, and E1BTF4, respectively. No tertiary structures were found in the deposited protein databank to cover the full studied proteins. Thus, the 3D structure of the involved *OVAL*, *OVALX*, and *OVALY* proteins were generated using the PhyRe2 (Protein Homology/analogY Recognition Engine) software (Kelley *et al*. 2015). The generated models were validated using Ramachandran plot (Laskowski *et al*. 2001) and QMEAN (Qualitative Model Energy ANalysis) Z-score generated by Swiss model structure assessment validation tools (Waterhouse *et al*. 2018). Protein Data Bank formats of the generated models were used as a template for predictions.

SNPs retrieval

All the amino acid substitutions of the three ovalbumin genes of chickens were comprehensively investigated in this study, namely *OVAL* (gene ID: 396058), *OVALX* (gene ID: 420898), and *OVALY* (gene ID: 420897) genes. A total of 30 nsSNPs were retrieved were retrieved from ensemble genome browser 97 (https://asia.ensembl.org/index.html), including 6 nsSNPs for *OVAL*, 15 nsSNPs for *OVALX*, and 9 nsSNPs for *OVALY* genes.

Prediction of nsSNPs effects on proteins structure, function, and stability

All nsSNPs were prioritized in terms of their effects on protein structure, function, and stability using ten different *in silico* tools. SIFT (Sorting Intolerant from Tolerant SNPs) (Pauline *et al*. 2003), PROVEAN (Protein Variation Effect Analyzer), Predict SNP (Bendl *et al*. 2014), MAPP (Multivariate Analysis of Protein Polymorphism) (Chao *et al*. 2008), PolyPhen (Polymorphism Phenotyping) (Adzhubei *et al*. 2013), and SNAP2 (Screening for Non-Acceptable Polymorphisms 2) (Smigielski *et al*. 2000) were utilized to assess the effect of each nsSNP on protein structure and biological activity. Meanwhile, mCSM (Pires *et al*. 2012), SDM (Worth *et al*. 2012), I-Mutant2 (Capriotti *et al*. 2005), and CUPSAT (Parthiban *et al*. 2006) tools were used to evaluate the effects of the same nsSNPs on proteins stability upon mutations.

Prediction of ligand binding sites of the most deleterious nsSNPs

The potential effect of each nsSNP on altering protein binding capability was predicted using set of protein-receptors or protein-ligand binding tools, including COACH (Wu *et al*. 2018), TM-SITE (Zhang *et al*. 2005), S-SITE (Yang *et al*. 2013), COFACTOR (Zhang *et al*. 2017), FINSIDTE (Brylinski *et al*. 2008), and ConCavity (Capra *et al*. 2009) prediction tools.

Energy analyses of proteins after being mutated with deleterious nsSNPs

Two sets of parameters were utilized to assess energy changes upon mutation. The total energy after minimization parameters and RMSD (root mean square deviation) values were calculated by DeepView/Swiss-PdbViewer (Johansson *et al*. 2012) and YASARA (Krieger *et al*. 2014) tools, respectively.

Results

The generation of optimal tertiary models of *OVAL* family genes was performed via PhyRe2 server. The validation of the generated models that carried out by PROCHECK and QMEAN tools indicated that the generated models were being within the accepted limits (Table 1). Considering the Ramachandran plot, models were considered qualified when they showed 90 % or more within the favoured regions. Results showed 91.3 % of the amino acid residues of *OVALX* and *OVALY* were suited in the favoured regions, while the amino acid residues of *OVAL* exhibited 90.0 % in the same region. Likewise, both *OVALX* and *OVALY* did not show any amino acid residues in the disallowed region, while *OVAL* showed only 2 amino acid residues (0.6 %) in the same region. QMEAN Z-score outputs were further proven the current models by giving -2.60, -2.20, and -1.92 for *OVALX*, *OVALY*, and *OVAL*, which, however, they were under the accepted standards.

Table 1. Three-dimensional structures validations for the PhyRe2 generated models of *OVALY*, *OVALX*, and *OVAL* proteins in chickens using PROCHECK tool and Swiss model assessments server.

Characteristics	***OVALX***	***OVALY***	***OVAL***
Residues in most favoured regions	324 (91.3 %)	324 (91.3 %)	320 (90.9 %)
Residues in additional allowed regions	27 (7.6 %)	27 (7.6 %)	29 (8.2 %)
Residues in generously allowed regions	4 (1.1 %)	4 (1.1 %)	1 (0.3 %)
Residues in disallowed regions	0 (0.0 %)	0 (0.0 %)	2 (0.6 %)
Number of non-glycine and non-proline residues	355 (100 %)	355 (100 %)	352 (100 %)
Number of end-residues	1	1	1
Number of glycine residues	20	20	19
Number of proline residues	12	12	14
Total number of residues	388	388	386
QMEAN Z-score	-2.60	-2.20	-1.92
Cβ	-1.49	-1.51	-0.77
All atoms	-2.32	-1.92	-1.60
Solvation	0.23	0.48	0.53
Torsion	-2.47	-2.15	-2.02

Table 2. Cumulative prediction for the most deleterious missense SNP in the OVALX, OVALY, and OVAL proteins in chickens using several *in silico* computation tools. The letters "D" refers to "deleterious", "N" refers to "Neutral", and "P" refers to "Prediction".

No.	SNP ID	AA substi-tution	SIFT	P	PROVEAN	P	Predict SNP	P	MAPP	P	PolyPhen	P	SNAP2	P	mCSM	P	SDM	P	I-Mutant2	P	CUPSAT	P
***A) OVALX* gene**																						
1.	rs741284084	E204Q	0.88	N	0.27	N	83%	N	0.74	N	0.67	N	-93	N	-0.206	D	-0.15	D	-1.12	D	-1.4	D
2.	rs314399376	G252D	0.7	N	-3.72	D	75%	N	0.71	N	0.67	N	29	D	-1.304	D	0.16	N	0.29	N	1.58	N
3.	rs315395834	S355F	0.69	N	-1.75	N	75%	N	0.77	N	0.67	N	-61	N	-0.551	D	0.98	N	-1.54	D	1.4	N
4.	rs13710183	L358F	0.4	N	-0.35	N	83%	N	0,7	N	0.67	N	-94	N	-0.606	D	-38	D	-0.04	D	-2.22	D
5.	rs13710184	L358S	0.07	N	0.25	N	83%	N	0.77	N	0.67	N	-83	N	-0.387	D	-0.61	D	-0.39	D	0.2	N
6.	rs738624565	**H365P**	**0**	**D**	**-8.9**	**D**	**87%**	**D**	**0.86**	**D**	**0.74**	**D**	**91**	**D**	**-0.924**	**D**	**-1.78**	**D**	**-1.56**	**D**	**-4.58**	**D**
B) *OVALY* gene																						
1.	rs3137354	V6L	0.04	D	-1.006	N	83%	N	0.71	N	0.67	N	-79	N	-0.837	D	0.68	N	-0.91	D	-0.57	D
2.	rs735157060	V80I	1	N	-0.131	N	83%	N	0.85	N	0.67	N	-93	N	0.092	N	-0.15	D	-1.15	D	2.8	N
3.	rs15113610	R165S	0.21	N	2.207	N	83%	N	0.74	N	0.67	N	-13	N	-1.718	D	-1.43	D	-2.49	D	0.32	N
4.	rs1057687043	**I167T**	**0.04**	**D**	**-3.373**	**D**	**72%**	**D**	**0.57**	**D**	**0.59**	**D**	**37**	**D**	**-2.974**	**D**	**-2.47**	**D**	**-2.48**	**D**	**-3.65**	**D**
5.	rs1059569127	F169L	0.51	N	0.387	N	74%	N	0.43	D	0.67	N	-66	N	-0.479	D	0.1	N	-3.02	D	2.15	N
6.	rs1058086850	G170D	41	N	1.062	N	75%	N	0.77	N	0.67	N	-65	N	-0.744	D	-0.46	D	-1.34	D	0.18	N
7.	rs1059731444	M173L	1	N	0.425	N	83%	N	0.79	N	0.67	N	-79	N	0.083	N	0.29	N	-1.42	D	3.36	N
8.	rs1060009844	F175L	1	N	3.216	N	74%	N	0.80	N	0.59	D	-77	N	-1.461	D	-1.89	D	-2.34	D	-1.89	D
9.	rs1057927044	I176V	1	N	0.900	N	83%	N	0.85	N	0.67	N	-90	N	-1.554	D	-2.15	D	-0.75	D	-1.77	D
10.	rs1060136897	T178A	1	N	3.405	N	83%	N	0.80	N	0.67	N	-55	N	-1.675	D	0.14	N	-1.68	D	0.78	N
11.	rs1057957612	I184M	0.15	N	1.301	N	83%	N	0.76	N	0.67	N	-92	N	-0.663	D	-0.9	D	-0.39	D	-1.58	D
12.	rs794348506	R195Q	0.85	N	0.493	N	83%	N	0.78	N	0.67	N	-88	N	0.031	N	-0.22	D	-1.04	D	-1.38	D
13.	rs16030775	E204Q	0.76	N	0.204	N	83%	N	0.75	N	0.67	N	-49	N	-0.213	D	-0.15	D	-0.10	D	-2.21	D
14.	rs1059306643	Y235F	0.07	N	-3.424	D	61%	D	0.76	N	0.67	N	-28	N	-1.002	D	-0.12	D	0.16	N	0.72	N
15.	rs731649580	Q255R	0.07	N	-1.272	N	83%	N	0.70	N	0.67	N	-74	N	-0.168	D	-0.23	D	-1.04	D	1.43	N
C) *OVAL* gene																						
1.	rs317913484	A6V	-60	N	-1.43	N	65%	N	0.64	N	0.67	N	-60	N	-0.429	D	-0.84	D	-1.18	D	-0.87	D
2.	rs733410745	**V209G**	**85**	**D**	**-6.77**	**D**	**87%**	**D**	**0.59**	**D**	**0.74**	**D**	**85**	**D**	**-2.465**	**D**	**-2.26**	**D**	**-1.44**	**D**	**-2.18**	**D**
3.	rs740916583	Y213S	31	D	-1.12	N	83%	N	0.76	N	0.67	N	31	D	-2.694	D	-1.25	D	-1.07	D	5.45	N
4.	rs734546333	**L231P**	**81**	**D**	**-6.39**	**D**	**87%**	**D**	**0.86**	**D**	**0.74**	**D**	**81**	**D**	**-1.525**	**D**	**-4.01**	**D**	**-0.14**	**D**	**-11.81**	**D**
5.	rs731320904	**F307C**	**68**	**D**	**-7.56**	**D**	**87%**	**D**	**0.86**	**D**	**0.74**	**D**	**86**	**D**	**-1.464**	**D**	**-1.05**	**D**	**-1.33**	**D**	**-2.81**	**D**
6.	rs739825984	N312D	-72	N	0.503	N	83%	N	0.85	N	0.67	N	-72	N	-1.374	D	-0.21	D	-0.47	D	-0.45	D
7.	rs733774347	**S317P**	**16**	**D**	**-4.14**	**D**	**76%**	**D**	**0.57**	**D**	**0.59**	**D**	**16**	**D**	**-0.537**	**D**	**-2.16**	**D**	**-1.40**	**D**	**-9.56**	**D**
8.	rs733242485	V342G	6	D	-0.69	N	63%	N	0.59	D	0.67	N	6	D	-0.548	D	-0.46	D	-1.23	D	-1.64	D
9.	rs734377844	V350G	80	D	-2.51	D	68%	N	0.46	D	0.67	N	80	D	-0.358	D	-0.72	D	-1.86	D	0.07	N

Table 3. Predictions of ligand binding sites for the most deleterious missense SNPs in the *OVALX*, *OVALY*, and *OVAL* proteins in chickens.

Protein Name	SNP ID	AA Change	COACH	TM-SITE	S-SITE	CO-FACTOR	FINDSITE	ConCavity
OVALX	rs738624565	*H365P*	Peptide (pdb id: 4au2D)	–	Multiple ligands	–	–	–
OVALY	rs1057687043	*I167T*	–	–	–	–	Ligand (pdb id: 1BR8B00)	–
OVAL	rs733410745	*V209G*	Peptide (pdb id: 3dy0A)	Ligand (1jrrA_BS01)	–	–	–	–
	rs734546333	*L231P*	Peptide (pdb id: 3dy0A)	Ligand (1lq8E_BS01)	–	Ligand (pdb id: 1iz2A)	–	–
	rs731320904	*F307C*	–	–	–	–	–	Binding residue
	rs733774347	*S317P*	–	–	Multiple ligands	–	Ligand (pdb id: 1AZXA00)	–

In order to evaluate the effect of each nsSNP on protein structure, biological activity, and stability, a variety of computational tools were utilized, including SIFT, PROVEAN, Predict SNP, MAPP, PolyPhen, SNAP2, mCSM, SDM, I-Mutant2, and CUPSAT. Cumulative results showed an entirely deleterious effect for a total six nsSNPs for the ovalbumin gene family, including *H365P* (rs738624565) for the *OVALX* gene, *I167T* (rs1057687043) for the *OVALY* gene, and *V209G* (rs733410745), *L231P* (rs734546333), *F307C* (rs731320904), and *S317P* (rs733774347) for the *OVAL* gene (Table 2).

More analyses were conducted on these most deleterious nsSNPs to explore their role to cause such extreme alteration in the mutant proteins. The ability of these risky nsSNPs to change the binding activity with other proteins was investigated by utilizing six prediction tools, including COACH, TM-SITE, S-SITE, COFACTOR, FINDSITE, and ConCavity. All the utilized prediction tools revealed obvious participation of all the included risky nsSNPs in several anticipated intensities with respect to their binding with ligands or receptors (Table 3). These findings may be due to the critical positioning taken by most of these nsSNPs in their proteins. Considering *H365P* and *I167T*, their positioning in the α-helix of the *OVALX* and *OVALY*, respectively may be involved in their deleterious effects (Fig. 1 A and B).

The 3D visualizations of the normal and mutant *OVAL* proteins were added another layer of confirmation of such deleterious effects caused by *V209G*, *L231P*, *F307C*, and *S317P*. This was due to their positioning within variable locations in the α-helix (Fig. 2).

The total energy values for the native *OVALX*, *OVALY*, and *OVAL* structures and their mutant structures were performed. Both *H365P* and *I167T*

Table 4. The total energy after minimization parameters and RMSD values for wild type and the most deleterious mutant forms of *OVALY*, *OVALX*, and *OVAL* proteins in chickens. The total energy after minimization and RMSD values were calculated by DeepView/Swiss-PdbViewer and YASARA tools, respectively.

Protein name	NsSNP ID	Amino acid change	Total energy after minimization	RMSD
OVALX	Wild type	-	-15699.480	-
	rs738624565	*H365P*	-15431.578	0.0219 A
OVALY	Wild type	-	-13697.646	-
	rs1057687043	*I167T*	-13707.831	0.0068 A
OVAL	Wild type	-	-17478.184	-
	rs733410745	*V209G*	-17381.457	0.0218 A
	rs734546333	*L231P*	-17443.225	0.0085 A
	rs731320904	*F307C*	-15665.894	0.0309 A
	rs733774347	S317P	-15635.917	0.0095 A

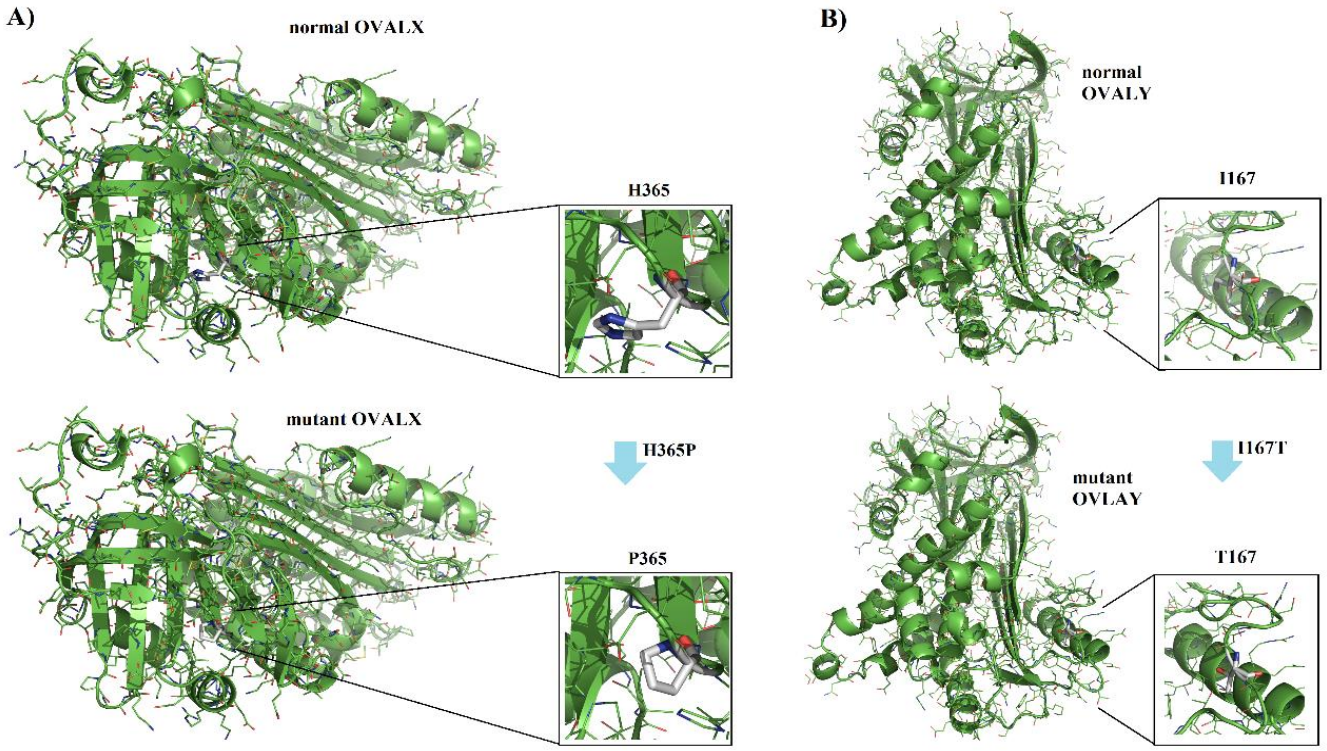

Fig. 1. Three-dimensional structure of the most deleterious nonsynonymous SNPs in the *OVALX* and *OVALY* proteins, namely *I90M*, *F109S*, and *Y127F*. **A.** positioning of the amino acid substitution *H365P* in the *OVALX* protein. **B**. positioning of the amino acid substitution *I167T* in the *OVALY* protein. The generated structures were visualized by PyMol software.

did not show a noticeable deviation in the total energy minimization values compared with their wild type counterparts. Considering *OVAL* deleterious mutations, both *F307C* and *S317P* mutant models showed an increase in energy minimization than the native structure (Table 4). These less favourable changes indicated more deleterious nature of both *F307C* and *S317P* models than the *V209G* and *L231P* counterpart.

Discussion

Ovalbumin family are the most essential proteins in hens' reproductive systems and are an ideal source for embryo development. Though several data have increasingly been provided with regard to *OVAL*, *OVALX*, and *OVALY* DNA sequences and its translation (Ren *et al*. 2017; DaSilva *et al*. 2019), no comprehensive study has been conducted to analyze the consequences of their nsSNPs on ovalbumin family. In this study, the retrieved nsSNPs of the ovalbumin gene family were comprehensively highlighted in terms of their effects on structure, biological activity, and stability. The present study witnessed six entirely deleterious nsSNPs in the ovalbumin gene family by recruiting up to 10 different *in silico* tools. However, all of these deleterious amino acid substitutions were found to be localized within several positions in the α-helix motif of *OVALX*, *OVALY*, and *OVAL* proteins. This localization may reflect a similar pattern of these substituted amino acid residues to induce these deleterious effects (Abrusán and Marsh 2016). In order to explore the possible roles of the observed nsSNPs in such deleterious effects, several *in silico* algorithms were used to perform this task. Moreover, the possible mechanisms of these nsSNPs in the interruption of these mutant proteins with regard to its binding with receptors were analyzed. Noticeable positioning of all these risky nsSNPs in binding sites with their corresponding receptors was notified. These findings indicated that the observed amino acid substitutions were involved in the binding network of the *OVALX*, *OVALY*, and *OVAL* protein, which may provide a clear modification in the affinity of interaction with corresponding receptors. Whatever the pattern each nsSNP takes in the interruption of the ovalbumin family, the current finding

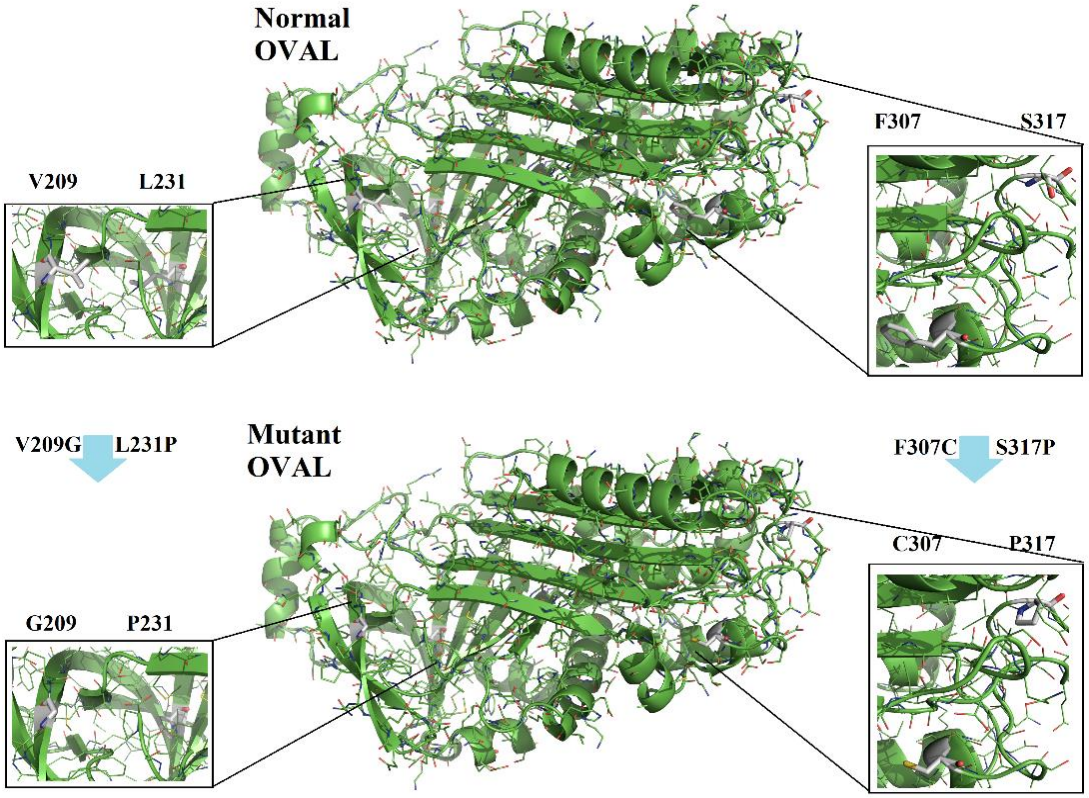

Fig. 2. Three-dimensional structure of the most deleterious nonsynonymous SNPs in the *OVAL* protein, namely *V209G*, *L231P*, *F307C*, and *S317P*. The generated structures were visualized by PyMol software.

indicated that these nsSNPs seriously affect the structure, function, and stability of ovalbumin family, which have a series of deleterious consequences on the biological pathways in which these proteins involved, such as egg defense activity, embryonic fertilization and development (Liu *et al.* 2015; Akazawa *et al.* 2019). Accordingly, its noteworthy to distinguish that both *H365P* and *I167T*, being the most deleterious nsSNPs in *OVALX* and *OVALY*, respectively, should take priority in terms of reducing antimicrobial activity in case of *OVALX* (Guyot *et al.* 2016), or damaging embryonic development in case of *OVALY* (Da Silva *et al.* 2015). This observation is particularly interesting, knowing that once these amino acid substitutions detected in a particular breed, serious precautions should be taken for breeders with respect to the handling of this breed. This is due to the presence of these SNPs in *OVALX* and its *OVALY* homolog, which may hamper the eggs production in this mutant population. The same precautions could also be applied to the *V209G*, *L231P*, *F307C*, and *S317P* mutant *OVAL* proteins. However, these amino acid substitutions deserve more attention as embryo development and the synthesis of many egg components are sensitive to alterations in *OVAL* architecture (Lin *et al.* 2016). Keeping in mind the wide availability of these proteins in all the eggs compartments (Dombre *et al.* 2017), the altered structure, biological activity, stability, or binding characteristics of ovalbumin proteins family seem to disrupt several biological pathways undertaken by chicks' embryos to take their schedules tasks of development and proper formation. It can be stated that the following nsSNPs *H365P*, *I167T*, *V209G*, *L231P*, *F307C*, and *S317P* should be absent in ovalbumin family in any chicken populations desired to be raised for higher egg production purposes.

Conclusions

The present cumulative *in silico* approach indicated the presence of entirely deleterious six nsSNPs in ovalbumin family, including *H365P* in *OVALX*, *I167T* in *OVALY*, and *V209G*, *L231P*, *F307C*, and *S317P* in *OVAL*. As long as more deleterious nsSNPs were witnessed in a particular breed, more serious indication for the reduced eligibility of this

breed for egg-production purposes should be concluded. This study provides the first comprehensive *in silico* guide for breeders to investigate the competency of raising chicken breeds for the large-scale egg production purposes.

Acknowledgement

This study has not received any fund from any agency or institution.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Abrusán G, Marsh JA (2016) Alpha helices are more robust to mutations than beta strands. PLoS Comput. Biol. 12: e1005242.

Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. 0 7: Unit 7.20.

Akazawa T, Ogawa M, Hayakawa S (2019) Migration of chicken egg-white protein ovalbumin-related protein X and its alteration in heparin-binding affinity during embryogenesis of fertilized egg. Poult. Sci. 98: 5100-5108.

Al-Shuhaib MBS, Al-Kafajy FR, Badi MA, AbdulAzeez S, Marimuthu K, Al-Juhaishi HAI, Borgio JF (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.

Al-Shuhaib MBS, Al-Kafajy FR., Al-Jashami GS (2019) A computational approach for explaining the effect of the *prl* gene polymorphism on prolactin structure and biological activity in Japanese quails. Anim. Biotechnol. 29: 1-9.

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput. Biol. 10: e1003440.

Brylinski M, Skolnick J (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc. Natl. Acad. Sci. USA 105: 129-134.

Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput. Biol. 5: e1000585.

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant 2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33: W306-W310.

Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H, Lynch H, Lipkin SM (2008) Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). Hum. Mutat. 29: 852-860.

Da Silva M, Beauclercq S, Harichaux G, Labas V, Guyot N, Gautron J, Nys Y, Rehault-Godbert S (2015) The family secrets of avian egg-specific ovalbumin and its related proteins Y and X. Biol. Reprod. 93: 71.

Da Silva M, Dombre C, Brionne A, Monget P, Chesse M, De Pauw M, Mills M, Combes-Soia L, Labas V, Guyot N, Nys Y, Réhault-Godbert S (2018) The unique features of proteins depicting the chicken amniotic fluid. Mol. Cell. Proteom. 18(Suppl 1): S174-S190.

Dombre C, Guyot N, Moreau T, Monget P, Da Silva M, Gautron J, Re´hault-Godbert S (2017) Egg serpins: the chicken and/or the egg dilemma. Semin. Cell Dev. Biol. 62: 120-132.

Guyot N, Labas V, Harichaux G, Chessé M, Poirier JC, Nys Y, Réhault-Godbert S (2016) Proteomic analysis of egg white heparin-binding proteins: towards the identification of natural antibacterial molecules. Sci. Rep. 6: 27974.

Hashim HO, Al-Shuhaib MBS, Ewadh MJ (2019) Heterogeneity of proteins in birds' egg-whites. Biotropia 26: 65-81.

Johansson MU, Zoete V, Michielin O, Guex N (2012) Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics 13: 173.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10: 845-858.

Krieger E, Vriend G (2014) YASARA View – molecular graphics for all devices – from smartphones to workstations. Bioinformatics 30: 2981-2982.

Laskowski RA, MacArthur MW, Thornton JM (2001) PROCHECK: validation of protein structure coordinates. *In* Rossmann MG and Arnold E (Ed.), International Tables of Crystallography, vol. F, Crystallography of Biological Macromolecules, Kluwer, Dordrecht, p. 722-725.

Lin YT, Wu CT, Huang JL, Cheng JH, Yeh KW (2016) Correlation of ovalbumin of egg white components with allergic diseases in children. J. Microbiol. Immunol. Infect. 49: 112-118.

Liu Y, Qiu N, Ma M (2015) Comparative proteomic analysis of egg white proteins during the rapid embryonic growth period by combinatorial peptide ligand libraries. Poult. Sci. 94: 2495-2505.

Mann K, Mann M (2008) The chicken egg yolk plasma and granule proteomes. Proteomics 8: 178-191.

Parthiban V, Gromiha MM, Schomburg D (2006) CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res. 34: W239-W242.

Pauline CN, Steven H (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31: 3812-3814.

Pires DEV, Ascher DB, Blundell TL (2012) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30: 335-342.

Qiu N, Liu W, Ma M, Zhao L, Li Y (2013) Differences between fertilized and unfertilized chicken egg white proteins revealed by 2-dimensional gel electrophoresis-based proteomic analysis. Poult. Sci. 92: 782-786.

Qiu N, Ma M, Cai Z, Jin Y, Huang X, Huang Q, Sun S (2012) Proteomic analysis of egg white proteins during the early phase of embryonic development. J. Proteomics 75: 1895-1905.

Rehault-Godbert S, Labas V, Helloin E, Herve-Grepinet V, Slugocki C, Berges M, Bourin MC, Brionne A, Poirier JC, Gautron J, Coste F, Nys Y (2013) Ovalbumin-related protein X is a heparin-binding ov-serpin exhibiting antimicrobial activities. J. Biol. Chem. 288: 17285-17295.

Rehault-Godbert S, Guyot N (2018) Vitellogenesis and yolk proteins, birds. *In* Skinner, M. (Ed.), Encyclopedia of Reproduction, Elsevier, Amsterdam, The Netherlands, London, UK, vol. 6, p. 3868.

Ren J, Hu J, Chen L, Liu Y, Xu X, He J, Shen J, Lu L (2017) Identification of egg white proteins and divergence in the regulatory region of the ovalbumin gene in avians. Protein Pept. Lett. 24: 12-25.

Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 28: 352-355.

Sugimoto Y, Sanuki S, Ibrahim HR, Aoki T, Kusakabe T, Koga K (2001) Occurrence of ovalbumin in ovarian yolk of chicken during oogenesis. Biochim. Biophys. Acta 1526: 1-4.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46: W296-W303.

Worth CL, Preissner R, Blundell TL (2012) SDM – a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res. 39: W215-W222.

Wu Q, Peng Z, Zhang Y, Yang J (2018) COACH-D: improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. Nucleic Acids Res. 46: W438-W442.

Yakubu1 A, De Donato M, Imumorin IG (2017) Modelling functional and structural impact of non-synonymous single nucleotide polymorphisms of the *DQA1* gene of three Nigerian goat breeds. S. Afr. J. Anim. Sci. 47: 146-156.

Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29: 2588-2595.

Zhang C, Freddolino PL, Zhang Y (2017) COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. 45: W291-W299.

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 22, 33: 2302-9.

Zhao JP, Zhang Q, Jiao HC, Wang XJ, Jiang MJ, Luo H, Lin H (2016) Ovalbumin expression in the oviduct magnum of hens is related to the rate of egg laying and shows distinct stress-type specific responses. J. Anim. Physiol. Anim. Nutr. 100: 876-883.