ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12880

Structural and dynamic analysis of G558R mutation in chicken *TSHR* gene shows altered signal transduction and corroborates its role as a domestication gene

A. Grottesi*, F. Gabbianelli†, A. Valentini† and G. Chillemi†,‡

*SCAI-Super Computing Applications and Innovation Department, CINECA, Via dei Tizii, 6b, 00185, Rome, Italy. †Department for Innovation in Biological, Agro-food and Forest systems, DIBAF, University of Tuscia, via S. Camillo de Lellis s.n.c., 01100, Viterbo, Italy. ‡Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, IBIOM, CNR, Via Giovanni Amendola, 122/O, Bari 70126, Italy.

Summary

The thyroid-stimulating hormone receptor (*TSHR*) has been indicated as a putative domestication gene in chicken. Comparison of WGS identified a variant in residue 558 of the transmembrane domain (TM) of TSHR, where the domestic chicken (GGD) presents an arginine, whereas the red jungle fowl (RJF) shares a conserved glycine with other vertebrates. This variant has been demonstrated to be associated with phenotypes that are important for domestication and related to thyroid regulation, such as less fearful behavior, reduced aggressive behavior and reduced dependence on seasonal reproduction in GGD as compared with RJF. By means of molecular dynamics simulations, we highlighted the structural and dynamic differences of variant Gly558Arg in the TSHR TM domain. Alterations in TM helix flexibility, structure and protein overall motion are described. The so-called 'arginine snorkeling' of residue 568 in GGD is observed and we hypothesize it as the originating force that produces the observed whole-protein perturbation in the helix bundle dynamics, capable of altering the TSHR signal transduction. The results are discussed in the context of their implications for a better understanding of biological mechanisms in chicken under control of the thyroid, such as body metabolism, as well as for their usefulness in biomedical research.

Keywords arginine snorkeling, chicken, domestication, G protein, molecular dynamics, thyroid, *TSHR*

Introduction

Domestication of the chicken (*Gallus gallus domesticus*, GGD) started around 8000 years ago in South and South-East Asia, with the double purpose of egg laying and meat production (Crawford, 1993; Tixier-Boichard *et al.*, 2011). Of particular importance for domestication and exploitation are tameness and year-long reproduction. In this respect, interest has been piqued by the thyroid-stimulating hormone receptor (*TSHR*), one the G protein-coupled receptors (Vassart, 2004), as a putative domestication gene, linked to photoperiod and reproduction (Karlsson *et al.*, 2016, 2015).

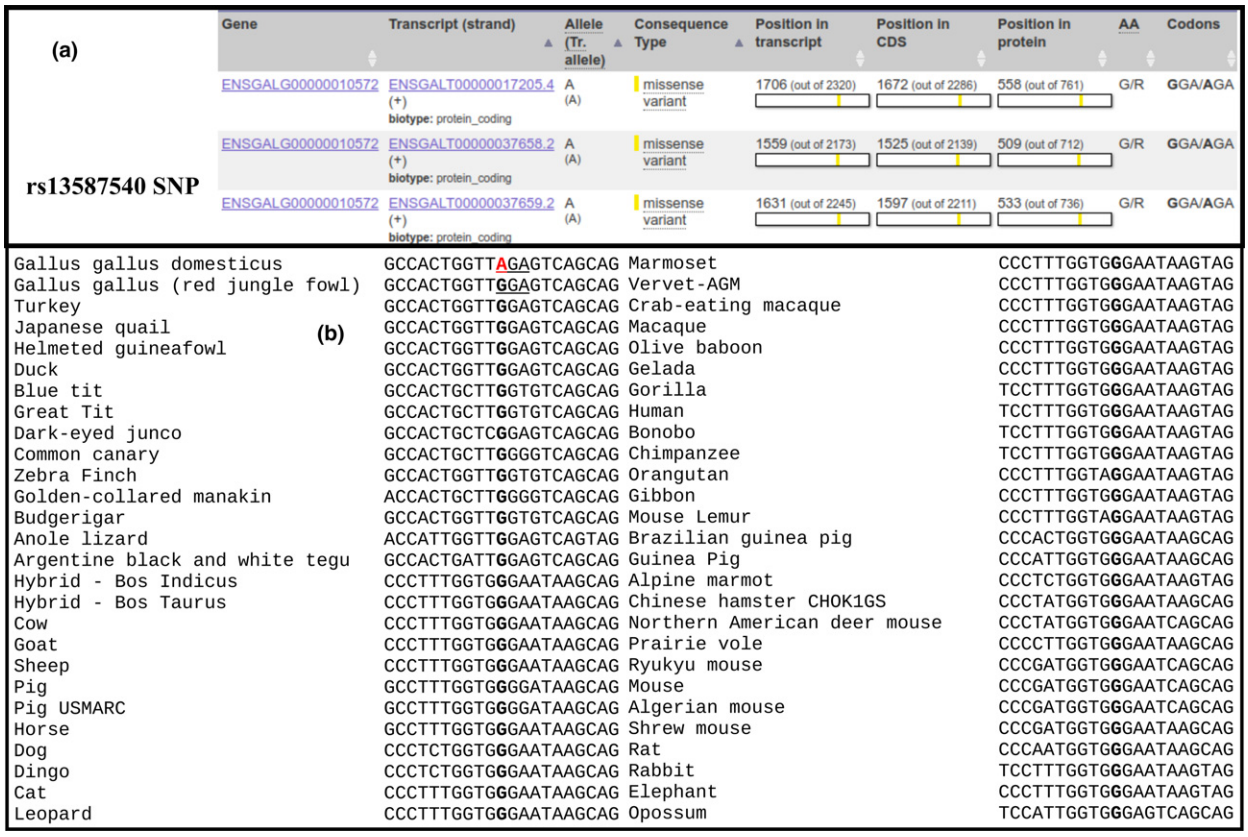
Address for correspondence

G. Chillemi, Department for Innovation in Biological, Agro-food and Forest systems, DIBAF, University of Tuscia, via S. Camillo de Lellis s.n.c., 01100 Viterbo, Italy.
E-mail: gchillemi@unitus.it

Accepted for publication 18 October 2019

A whole-genome comparison between domestic chicken and its major wild ancestor, the red jungle fowl (*Gallus gallus*, RJF), identified the Gly558Arg variant in the *TSHR* as a possible 'domestication locus' (Rubin *et al.*, 2010). The AGA codon in the *TSHR* TM sequence, encoding the arginine residue (rs13587540 SNP, reported in Fig. 1), in fact, is observed in GGD whereas the RJF shares the highly conserved guanine with the 54 amniota vertebrates Mercator-Pecan (for more details on this multi-alignment, see https://www.ensembl.org/info/genome/compara/mlss.html?mlss=1546) and has the glycine residue in position 558.

Phenotypic characterization of this *TSHR* variant has been investigated in offspring of the F8 generation of an intercross between White Leghorn GGD and RJF (Karlsson *et al.*, 2015). Homozygous chickens showed longer incubation times, less fearful behaviors, lower numbers of aggressive behaviors and decreased levels of the thyroid hormone when carrying the domestic instead of the wild genotype (Karlsson *et al.*, 2015).



(a)

Gene	Transcript (strand)	Allele (Tr. allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	AA	Codons
ENSGALG00000010572	ENSGALT00000017205.4 (+) biotype: protein_coding	A (A)	missense variant	1706 (out of 2320)	1672 (out of 2286)	558 (out of 761)	G/R	GGA/AGA
ENSGALG00000010572	ENSGALT00000037658.2 (+) biotype: protein_coding	A (A)	missense variant	1559 (out of 2173)	1525 (out of 2139)	509 (out of 712)	G/R	GGA/AGA
ENSGALG00000010572	ENSGALT00000037659.2 (+) biotype: protein_coding	A (A)	missense variant	1631 (out of 2245)	1597 (out of 2211)	533 (out of 736)	G/R	GGA/AGA

rs13587540 SNP

(b)

Gallus gallus domesticus	GCCACTGGTTAGAGTCAGCAG	Marmoset	CCCTTTGGTGGGAATAAGTAG
Gallus gallus (red jungle fowl)	GCCACTGGTTGGAGTCAGCAG	Vervet-AGM	CCCTTTGGTGGGAATAAGTAG
Turkey	GCCACTGGTTGGAGTCAGCAG	Crab-eating macaque	CCCTTTGGTGGGAATAAGTAG
Japanese quail	GCCACTGGTTGGAGTCAGCAG	Macaque	CCCTTTGGTGGGAATAAGTAG
Helmeted guineafowl	GCCACTGGTTGGAGTCAGCAG	Olive baboon	CCCTTTGGTGGGAATAAGTAG
Duck	GCCACTGGTTGGAGTCAGCAG	Gelada	CCCTTTGGTGGGAATAAGTAG
Blue tit	GCCACTGCTTGGTGTCAGCAG	Gorilla	TCCTTTGGTGGGAATAAGTAG
Great Tit	GCCACTGCTTGGTGTCAGCAG	Human	TCCTTTGGTGGGAATAAGTAG
Dark-eyed junco	GCCACTGCTCGGAGTCAGCAG	Bonobo	TCCTTTGGTGGGAATAAGTAG
Common canary	GCCACTGCTTGGGGTCAGCAG	Chimpanzee	TCCTTTGGTGGGAATAAGTAG
Zebra Finch	GCCACTGGTTGGTGTCAGCAG	Orangutan	CCCTTTGGTAGGAATAAGTAG
Golden-collared manakin	ACCACTGCTTGGGGTCAGCAG	Gibbon	CCCTTTGGTGGGAATAAGTAG
Budgerigar	GCCACTGGTTGGTGTCAGCAG	Mouse Lemur	CCCTTTGGTAGGAATAAGTAG
Anole lizard	ACCATTGGTTGGAGTCAGTAG	Brazilian guinea pig	CCCACTGGTGGGAATAAGCAG
Argentine black and white tegu	GCCACTGATTGGAGTCAGCAG	Guinea Pig	CCCATTGGTGGGAATAAGCAG
Hybrid - Bos Indicus	CCCTTTGGTGGGAATAAGCAG	Alpine marmot	CCCTCTGGTGGGAATAAGTAG
Hybrid - Bos Taurus	CCCTTTGGTGGGAATAAGCAG	Chinese hamster CHOK1GS	CCCTATGGTGGGAATAAGCAG
Cow	CCCTTTGGTGGGAATAAGCAG	Northern American deer mouse	CCCTATGGTGGGAATCAGCAG
Goat	CCCTTTGGTGGGAATAAGCAG	Prairie vole	CCCCTTGGTGGGAATAAGCAG
Sheep	CCCTTTGGTGGGAATAAGCAG	Ryukyu mouse	CCCGATGGTGGGAATCAGCAG
Pig	GCCTTTGGTGGGGATAAGCAG	Mouse	CCCGATGGTGGGAATCAGCAG
Pig USMARC	GCCTTTGGTGGGGATAAGCAG	Algerian mouse	CCCGATGGTGGGAATCAGCAG
Horse	GCCTTTGGTGGGAATAAGCAG	Shrew mouse	CCCGATGGTGGGAATAAGCAG
Dog	CCCTCTGGTGGGAATAAGCAG	Rat	CCCAATGGTGGGAATAAGCAG
Dingo	CCCTCTGGTGGGAATAAGCAG	Rabbit	TCCTTTGGTGGGAATAAGTAG
Cat	CCCTTTGGTGGGAATAAGCAG	Elephant	CCCTTTGGTGGGAATAAGTAG
Leopard	CCCTTTGGTGGGAATAAGCAG	Opossum	TCCATTGGTGGGAGTCAGCAG

Figure 1 Relevant information for the rs13587540 variant in the *Gallus gallus* species (location chromosome 5:41020238; forward strand) from Ensembl databases and web tools (Hunt et al., 2018). (a) Information on the three *TSHR* isoforms containing the rs13587540 variant, obtained with the 'Genes and regulation' panel. We followed the residue numbering from the first transcript as in Girdland Flink et al. (2014). (b) Multiple alignment in '54 amniota vertebrates Mercator-Pecan' obtained by the 'Phylogenetic Context' panel. Only GGA shows an A base (red) whereas the other vertebrates have a G base in the same position (shown in bold). The corresponding codons are underlined in *G. gallus* species.

A further study of the same authors also demonstrated that this *TSHR* variant affects the photoperiodic response in GGD by reducing dependence on seasonal reproduction as compared with RJF (Karlsson *et al.*, 2016). Actually, RJF have been proposed as an animal model for studying the molecular mechanism of seasonal reproduction owing to their strong photoperiodic sensibility (Ono *et al.*, 2009).

Therefore, the Gly558Arg TSHR variant affects typical domestication traits and its importance during the evolution of the domestic chicken is also demonstrated by its presence already in birds present in the Roman Empire, even if the allele became fixed only 500 years ago, probably owing to human-driven selection (Girdland Flink *et al.*, 2014).

Here we report molecular dynamics (MD) simulations of the transmembrane (TM) domain of TSHR in a double lipid layer to characterize at molecular level the effect of the Gly558Arg variant on the flexibility and internal dynamics of TM helices, responsible for TSHR signal transduction between extra- and intracellular compartments (Vassart, 2004).

MD simulation of biological macromolecules is a powerful tool for examining structural and dynamic properties as it provides a description at atomic level and at the appropriate timescale (Biagini *et al.*, 2018). The perturbation effect of single-residue mutations, in particular, is efficiently highlighted by coupled MD simulations of wt and mutant proteins (Chillemi *et al.*, 2008; Bongiorni *et al.*, 2016).

An MD study of TSHR TM domain in *Homo sapiens* has been recently carried out (Chantreau *et al.*, 2015). Although limited in the total simulation time (60 ns), the study revealed structural determinants of the TM domain of TSHR and related receptors with the aim of rationally designing ligands targeting the receptor's internal cavity (Chantreau *et al.*, 2015). In fact, research on small-molecule TSHR agonists and antagonists is active, with the aim of cure numerous pathologies linked to thyroid malfunction (Davies & Latif, 2015; Smith, 2017).

Our MD study of the G558R mutation is aimed at understanding the molecular mechanism behind the single mutation that has allowed GGD to become such an important commodity worldwide and paving the way to discovery of the physiological features of other genetic variants in the same domain of *TSHR*.

Results

TSHR in the *Gallus gallus* species has three known isoforms (Fig. 1a) that differ in their *N*-terminal region. However, the TM domain, responsible for signal transduction, is completely conserved and only the residue numbering changes. We followed the numbering according to the first transcript as in (Girdland Flink *et al.,* 2014).

GGD is the only vertebrate with an AGA codon in the TSHR TM domain, derived from the mutation of its first base. All other vertebrates (including RJF), in fact, show mutations only in the third base of the codon. Therefore, the observed GGA, GGT and GGG codons are synonymous and all encode for the Glycine residue, whereas only GGD has an arginine residue in position 558 (Fig. 1b).

The structure of the TSHR TM domain is composed by seven TM helices (in the following named from TM1 to TM7, shown in Fig. 1, where Arg558 is highlighted in red. This TSHR domain is responsible for transduction of the signal after the *N*-terminal portion of the protein binds with high affinity the thyroid-stimulating hormone (TSH) (Smits *et al.,* 2002). Interaction of TSH/TSHR plays a role in physiological functions such as thyroid gland activity and regulation of metabolic cycles (Vassart, 2004; McNabb, 2007).

The overall fold stability of the TSHR TM domain was checked by calculating the protein root mean square deviations (RMSDs) of RJF and GGD variants from the respective starting structure and as a function of the simulation time (black and red respectively, in Figure S1). An unstable protein, in fact, will produce a continuously growing RMSD curve during the unfolding process.

In our case, both systems have stable folds, with an RMSD plateau of 5 and 6 Å for RJF and GGD respectively. Moreover, GGD showed a longer relaxation time (40 vs 25 ns in RJF), i.e. the simulation time with growing RMSD, suggesting a more complex structural rearrangement in this system. As we are interested in the equilibrium properties of the protein under physiological conditions, we discarded the first 50 ns of simulation from all of the following analyses in both systems.

The flexibility of the two proteins has been measured through root mean square fluctuation (RMSF) calculation (Fig. 2). GGD shows higher fluctuations, particularly in the inter-helix loops between TM1/TM2, TM3/TM4 and TM5/TM6, but also large portions of TM1, TM2 and nearly all TM4 show the same increased flexibility. On the contrary, the *C*-terminal portion of TM6 and the loop TM6/TM7 are less flexible in GGD as compared with RJF. It is worth noting that residue 558 (highlighted in Fig. 1 with a vertical arrow) has similar RMSF values in both systems, suggesting that this residue variant may have non-local effect on protein dynamics.

We then studied the protein global motion by means of essential dynamics (ED) (Amadei *et al.,* 1993). This analysis separates the main correlated motions, which are biologically relevant, from the irrelevant local fluctuations by

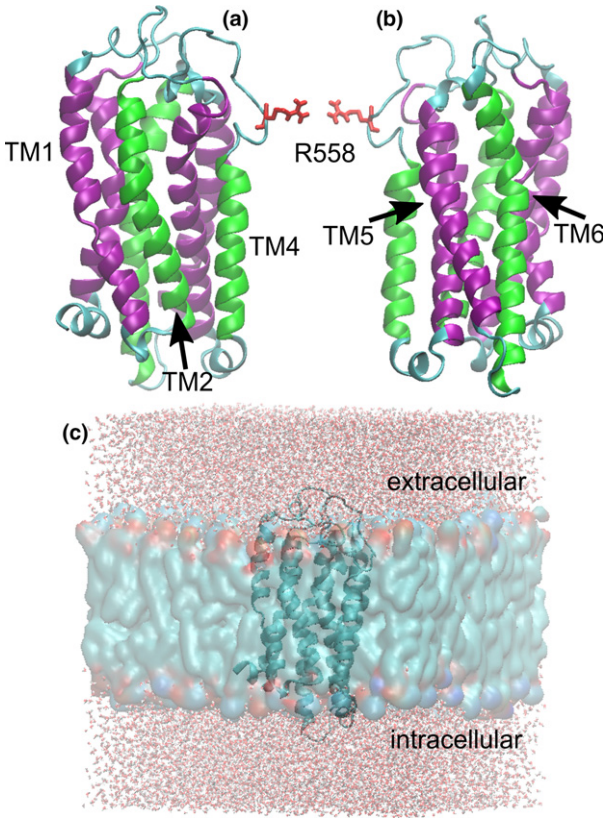


Figure 2 3D structure of the TSHR TM domain (starting structure for GGD). (a, b) Two 90° rotated orientations. Residue Arg558 is highlighted in red. Even- and odd-numbered TM helices are colored in green and purple respectively. (c) The full system of TSHR TM domain embedded in the lipid membrane bilayer and immersed in a box of explicit water molecules (starting structure).

projecting the total motions along principal eigenvectors. We first carried out an ED analysis on a concatenated trajectory of both systems (see Methods section). The 2D projection on the essential subspace along eigenvectors 1 and 2 shows that the two simulations sample two separate subspaces along eigenvector 1 (Fig. 3). The single mutation G558R, therefore, provokes a dramatic change in the protein global motion. Moreover, the GGD MD (red dots in Fig. 3) shows a more pronounced flexibility along the eigenvector 2, as evidenced by the extended conformational variability along this eigenvector as compared with RJF.

The different dynamics of the two systems were further investigated by plotting the bend angle for all TM helices (Fig. 1 and Figure S2). The three TM helices behaved differently in the two systems. TM2 and TM4, in particular, showed a greater bend angle in GGD with respect to RJF, whereas the contrary was observed for TM7 (Fig. 1). It is noteworthy that residue 558 is close to the *C*-terminal region of TM4. All of the other TM helices have comparable bend angle values in the two systems (Figure S2).

In order to highlight the different motions of the TM helices, we carried out a second ED analysis on the two

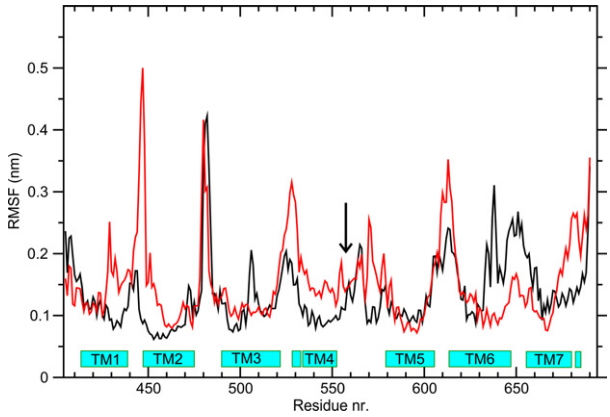


Figure 3 Per-residue RMSFs of RJF and GGD TSHR TM domain are reported in black and red respectively. A vertical arrow indicates the variant residue position along the protein chain.

separated systems. In fact, the first analysis, being on the concatenated trajectory, generated a single set of eigenvectors on which both trajectories were projected, thus highlighting the different visited subspaces. With this second ED analysis, differently, we aimed at generating two separate sets of eigenvectors to extract the most significant biological motions by each simulation and comparing them. Eigenvector fluctuations along eigenvectors 1–3 are shown in Figure S3. This analysis shows that the fluctuation along the first and third eigenvectors is similar in the two systems. In contrast, several differences, involving the whole protein domain, are shown in the projection along the second eigenvector that accounts for 12 and 13% of the total motion in GGD and RJF respectively. These results, therefore, suggest that this single component is capable of detecting nearly the whole of the differences in the movement of the TM helix bundle between GGD and RJF, responsible for TSHR signal transduction from extra- to intracellular compartments (Vassart & Dumont, 1992). The 3D animation of the projection of the two trajectories along the respective second eigenvector is shown as a movie in Supplementary material. The movie clearly shows that the dynamic properties of the TM, and therefore the TSHR signal transduction, are completely different in the two systems.

Protein–lipid interactions play an important role in TM domain dynamics and regulation of function for many membrane proteins (Laganowsky *et al.*, 2014). Here we checked the hypothesis that the observed alteration in TM dynamics could be related to an altered network of protein–lipid interactions, local to the variant residue. Therefore, we plotted the minimum distance between the 558 residue main chain atoms with the polar heads of lipids in both the GGD and RJF systems (Fig. 2a). As the figure clearly shows, this distance is greater in RJF compared with GGD (compare black vs red lines).

Moreover, the total number of protein–lipid contacts (Fig. 2b) is significantly higher in GGD than in RJF TSHR, suggesting that the conformational dynamics of the TM domain in RJF results in major exposure of the residue sidechains to the water–lipid interface. In particular, the Arg558 side-chain is located at the *N*-terminal of the TM4 helix, near the protein–lipid interface, making possible ionic interactions between the guanidinium moiety of the arginine side-chain and the polar head-groups of palmitoyl-oleoyl-phosphatidylcholine (POPC) lipids. This type of interaction, usually referred to as 'arginine snorkeling' (Keskin *et al.*, 2017), clearly provides evidence that Arg558 moves toward the membrane border, thus explaining the observed deformation in the TSHR TM helices of GGD and the alteration in TM dynamics.

The arginine snorkeling phenomenon in GGD is well visualized by a 3D representation of the protein structure at 300 ns, where the movement of Arg558 toward the polar head-groups of the lipid membrane is highlighted and accompanied by a strong kink in TM4 (Fig. 3a). On the contrary, the Glycine residue in RJF does not greatly change its relative position in the TM helix bundle and TM4 maintains its starting linear conformation (Fig. 3b). From the conservation of the fluctuations along the first and third eigenvectors, that together account for 37% of the total motion in both systems, we can safely hypothesize that the observed structural and dynamic differences in GGD as compared with RJF are not enough to completely abolish the TSHR signal transduction, but they produce a modification of the thyroid activity, in line with natural occurring mutations in the TM of human TSHR (Kleinau & Vassart, 2000).

Discussion

The Arg568Gly variant in *TSHR* affects typical domestication traits related to thyroid function (Karlsson *et al.*, 2016, 2015). As it has been demonstrated that the extracellular domain at the *N*-terminal region of TSHR is sufficient to bind TSH with high affinity (Da Costa & Johnstone, 1998), the Arg568Gly variant only affects the structural and dynamic behavior of the TM domain, responsible for signal transduction between the extra- and intracellular compartment (Vassart & Dumont, 1992).

We characterized the alteration in TM helix flexibility (Fig. 2), structure (Fig. 1) and protein–lipid interactions (Fig. 2) between GGD and RJF genotypes. By means of ED analysis, we extracted the eigenvector that contains nearly all of the dynamic differences in the two systems (Movie S1). Finally, we characterized the region harboring the variant and observed the long-range protein perturbation in GGD. The 'snorkeling' of Arg568 in GGD, i.e. the interaction of its lateral chain with the lipid polar chains, is clearly present in our simulation (Fig. 3). It is well known that the arginine residue is only tolerated in the membrane environment by the snorkeling of its guanidinium sidechain, which therefore escapes from the hydrophobic environment of the lipid tails (Keskin *et al.*, 2017; Ulmschneider *et al.*, 2017).

Gly568 in RJF, on the other hand, maintains its original position with no interaction with the lipid polar heads (Fig. 3). Finally, we observe a strong bend of the near TM4 helix in GDD, whereas it maintains its linearity in RJF (Fig. 3).

Our results demonstrate an important alteration in the TSHR signal transduction mechanism along the second eigenvector in the ED analysis, which accounts for 12 and 13% of the total motion in GGD and RJF. A greater portion of the total motion (35% in both systems) is nevertheless conserved in the two systems, as shown by the fluctuations along eigenvectors 1 and 3. We are able, therefore, to produce a coherent structural and dynamic interpretation for the observed phenotypic differences between RJF and GGD related to thyroid regulation (Karlsson *et al.*, 2016, 2015). As already found in human (Chantreau *et al.*, 2015), the helix dynamics of the TM domain comprise an important aspect of TSHR activation and our study may prove useful in the rational design of small molecules that regulate TSHR activation, with a possible impact also on medical research. Dysfunctions of the TSHR are indeed the underlying cause of various gain- or loss-of-function phenotypes associated with thyroid malfunction (Kleinau & Vassart, 2000).

The approach we have followed may also be considered in other studies in the livestock sector to explain the effect of the several mutations that have been discovered but are not yet characterized at molecular level and for which an explanation of their physiological consequences has not been achieved.

Materials and methods

Homology modeling

Initial structures of transmembrane domain of wild-type and domestic TSHR were determined by comparative homology modeling. A hierarchical approach was used to find the best representative 3D model structure using the chicken TSHR primary structure. We used the GPCR-I-TASSER web service (Zhang *et al.*, 2015), which is based on the I-TASSER (Iterative Threading ASSEmbly Refinement) program. It first identifies structural templates from the PDB by a multiple threading approach with full-length atomic models constructed by iterative template-based fragment assembly simulations. The procedure generated 10 models and the model with the lower *C*-Score and higher cluster density (−4.07 and 0.185 respectively) was chosen as the best representative for the final structure. Variant residue was introduced by replacing the residue side-chain with the corresponding one taken from a rotamer library, with the Pymol (www.pymol.org) software.

MD simulation protocol

MD simulations of wt and domestic TSHR were performed using the homology modeling-derived structures. All simulations were performed using gromacs 2018.x. The Charmm27 force field was used (Mackerell *et al.*, 2004). Initial structures were immersed in a triclinic simulation box, embedded in a POPC membrane bilayer and then solvated with TIP3P water molecules (Berendsen *et al.*, 1981). Ionic strength was adjusted so as to neutralize the total charge of the simulation box. All simulations were performed in the NVT ensemble at constant volume and constant temperature (300 K). Periodic boundary conditions were then applied. Initial velocities were taken from the Maxwell–Boltzmann distribution at 300 K. Long-range electrostatic interactions were calculated using the particle mesh Ewald method (Darden *et al.*, 1993) with a 1.2 nm cut-off for the real space calculation. A 1.2 nm cut-off was used to estimate Van der Waals interactions. The pair list was updated every 10 steps. The LINCS algorithm was used to constrain bond lengths (Hess *et al.*, 1997); the time step for

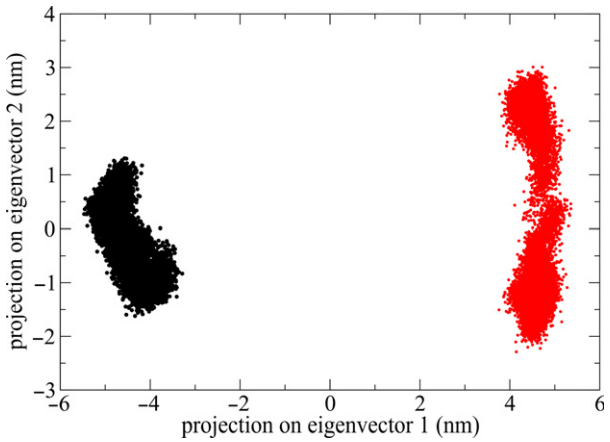


Figure 4 2D projection of RJF and GGD systems on the essential subspace along eigenvectors 1 and 2 of the concatenated MD trajectory (see Methods for details).

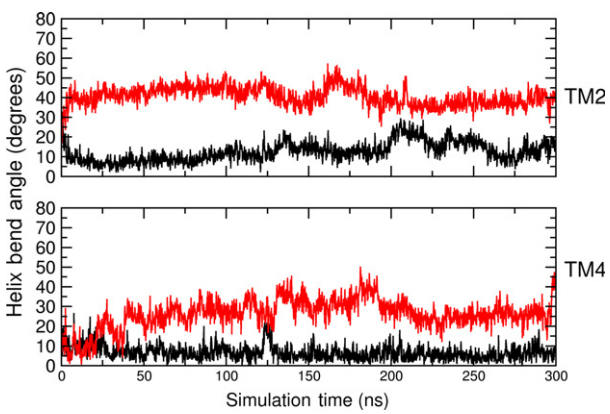


Figure 5 Maximum helix bend angle (in degrees) calculated according to Bendix (see Methods for details) for TM helices 2, 4, and 7 as a function of simulation time (in nanoseconds). For each panel the bend angle for GGD (red) and RJF (black) TSHR is shown.

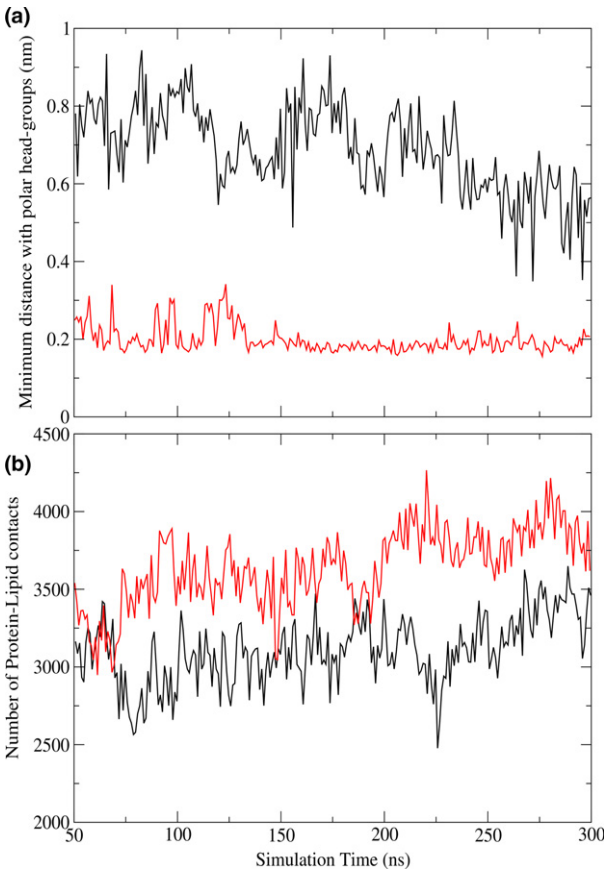


Figure 6 Protein–lipid interactions for RJF (black) and GGD (red) as a function of simulation time (in nanoseconds). (a) Minimum distance between the side-chain moiety of residue at position 558 and the polar head-groups of POPC phospolipids. (b) The total number of contacts (within 0.6 nm) between all protein side-chains and the polar head-groups of POPC phospolipids are shown.

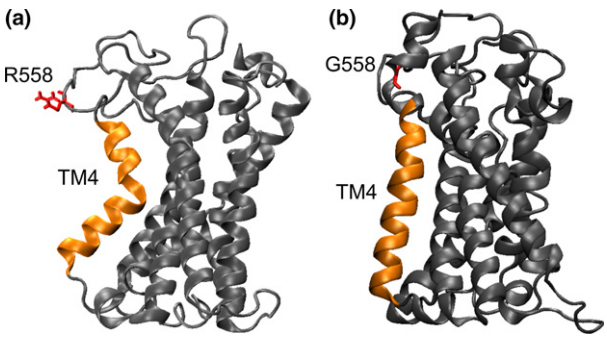


Figure 7 3D structures at 300 ns simulation time. (a) GGD system TSHR secondary structure is shown as dark gray cartoon; TM4 helix is highlighted in orange and Arg558 as red sticks. (b) Same as (a) for RJF. Glycine 558 is highlighted as red sticks.

integration was 2 fs. Initial structures for all simulations were subjected to a steepest-descent minimization cycle to reduce steric hindrance. Then a restrained MD step-wise procedure was applied to gradually release the restraints and allow the system to equilibrate at the simulated temperature of 300 K: the applied restrained were 1000, 500 and 250 kJ/mol. The total simulation time was 300 ns, typically. Coordinates were saved every 20 ps.

Analysis

Helix bending angles were measured using BendIx (Dahl *et al.*, 2012), a VMD (Humphrey *et al.*, 1996) extension that allows the dynamic measurement of the maximum bend angles along an MD simulation trajectory. Lipid–protein interactions were analyzed by measuring the number of contacts of residues side-chains with all lipid head-groups within a sphere of 0.6 nm radius. Lipid head-groups were defined as all heavy atoms of the choline moiety. Principal components analysis (essential dynamics) was used to analyze the large-amplitude motions of protein dynamics by projection of the MD simulations onto relevant eigenvectors (with largest eigenvalue) of the positional fluctuation covariance matrix (Amadei *et al.*, 1993). ED was performed on each individual MD trajectory as well as on the concatenated trajectory resulting from the merging of the single equilibrated trajectories (beyond 50 ns).

Acknowledgements

We acknowledge Cineca for computing resources. This research was supported by the 'Departments of Excellence-2018' Program (Dipartimenti di Eccellenza) of the Italian Ministry of Education, University and Research, DIBAF-Department of University of Tuscia, Project 'Landscape 4.0 – food, wellbeing and environment'.

References

Amadei A., Linssen A.B., Berendsen H.J. (1993) Essential dynamics of proteins. *Proteins* **17**, 412–25.

Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., Hermans J. (1981) Interaction Models for Water in Relation to Protein Hydration, in: Pullman B (Ed.), *Intermolecular Forces*. Springer Netherlands, Dordrecht, pp. 331–342.

Biagini T., Chillemi G., Mazzoccoli G., Grottesi A., Fusilli C., Capocefalo D., Castellana S., Vescovi A.L., Mazza T. (2018) Molecular dynamics recipes for genome research. *Briefings in Bioinformatics* **19**, 853–62.

Bongiorni S., Valentini A., Chillemi G. (2016) Structural and dynamic characterization of the C313Y mutation in myostatin dimeric protein, responsible for the 'Double Muscle' phenotype in Piedmontese cattle. *Frontiers in Genetics* **7**, 14, https://doi.org/10.3389/fgene.2016.00014.

Chantreau V., Taddese B., Munier M., Gourdin L., Henrion D., Rodien P., Chabbert M. (2015) Molecular insights into the transmembrane domain of the thyrotropin receptor. *PLoS ONE* **10**, e0142250.

Chillemi G., D'Annessa I., Fiorani P., Losasso C., Benedetti P., Desideri A. (2008) Thr729 in human topoisomerase I modulates

anti-cancer drug resistance by altering protein domain communications as suggested by molecular dynamics simulations. *Nucleic Acids Research* **36**, 5645–51.

Da Costa C.R., Johnstone A.P. (1998) Production of the thyrotrophin receptor extracellular domain as a Glycosylphosphatidylinositol-anchored Membrane Protein and Its Interaction with Thyrotrophin and Autoantibodies. *Journal of Biological Chemistry* **273**, 11874–80.

Crawford R.D. ed. (1993) *Poultry breeding and genetics, 2*, print edn. Amsterdam: Developments in animal and veterinary science. Elsevier.

Dahl A.C.E., Chavent M., Sansom M.S.P. (2012) Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* **28**, 2193–4. https://doi.org/10.1093/bioinformatics/bts357.

Darden T., York D., Pedersen L. (1993) Particle mesh Ewald: An *N* ·log(*N*) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–92. https://doi.org/10.1063/1.464397.

Davies TF, Latif R (2015) Targeting the thyroid-stimulating hormone receptor with small molecule ligands and antibodies. *Expert Opinion on Therapeutic Targets* **19**, 835–47.

Girdland Flink L., Allen R., Barnett R., Malmstrom H., Peters J., Eriksson J., Andersson L., Dobney K., Larson G. (2014) Establishing the validity of domestication genes using DNA from ancient chickens. *Proceedings of the National Academy of Sciences* **111**, 6184–9.

Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–72.

Humphrey W., Dalke A., Schulten K. (1996) VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**(33–38), 27–8.

Hunt S.E., McLaren W., Gil L. *et al*. (2018) Ensembl variation resources. *Database* **2018**, https://doi.org/10.1093/database/bay119.

Karlsson A.-C., Svemer F., Eriksson J., Darras V.M., Andersson L., Jensen P. (2015) The Effect of a Mutation in the Thyroid Stimulating Hormone Receptor (TSHR) on Development, Behaviour and TH Levels in Domesticated Chickens. *PLoS ONE* **10**, e0129040.

Karlsson A.-C., Fallahshahroudi A., Johnsen H., Hagenblad J., Wright D., Andersson L., Jensen P. (2016) A domestication related mutation in the thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and reproduction in chickens. *General and Comparative Endocrinology* **228**, 69–78.

Keskin A., Akdoğan E., Dunn C.D. (2017) Evidence for amino acid snorkeling from a high-resolution, *In Vivo* analysis of Fis1 Tail-Anchor insertion at the mitochondrial outer membrane. *Genetics* **205**, 691–705.

Kleinau G., Vassart G. (2000) TSH Receptor Mutations and Diseases. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, Hershman JM, Kaltsas G, Koch C, Kopp P, Korbonits M, McLachlan R, Morley JE, New M, Perreault L, Purnell J, Rebar R, Singer F, Trence DL, Vinik A, Wilson DP, eds. *Endotext*. South Dartmouth (MA): MDText.com Inc.

Laganowsky A., Reading E., Allison T.M., Ulmschneider M.B., Degiacomi M.T., Baldwin A.J., Robinson C.V. (2014) Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–5.

Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* **25**, 1400–15.

McNabb F.M.A. (2007) The hypothalamic-pituitary-thyroid (HPT) axis in birds and its role in bird development and reproduction. *Critical Reviews in Toxicology* **37**, 163–93. https://doi.org/10.1080/10408440601123552.

Ono H., Nakao N., Yamamura T., Kinoshita K., Mizutani M., Namikawa T., Iigo M., Ebihara S., Yoshimura T. (2009) Red jungle fowl (*Gallus gallus*) as a model for studying the molecular mechanism of seasonal reproduction. *Animal Science Journal* **80**, 328–32.

Rubin C.-J., Zody M.C., Eriksson J. *et al*. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591.

Smith T.J. (2017) TSHR as a therapeutic target in Graves' disease. *Expert Opinion on Therapeutic Targets* **21**, 427–32.

Smits G., Govaerts C., Nubourgh I., Pardo L., Vassart G., Costagliola S. (2002) Lysine 183 and Glutamic Acid 157 of the TSH Receptor: Two Interacting Residues with a Key Role in Determining Specificity toward TSH and Human CG. *Molecular Endocrinology* **16**, 722–35.

Tixier-Boichard M., Bed'hom B., Rognon X (2011) Chicken domestication: From archeology to genomics. *Comptes Rendus Biologies* **334**, 197–204.

Ulmschneider M.B., Ulmschneider J.P., Freites J.A., von Heijne G., Tobias D.J., White S.H. (2017) Transmembrane helices containing a charged arginine are thermodynamically stable. *European Biophysics Journal* **46**, 627–37.

Vassart G. (2004) A molecular dissection of the glycoprotein hormone receptors. *Trends in Biochemical Sciences* **29**, 119–26.

Vassart G., Dumont J.E. (1992) The thyrotropin receptor and the regulation of thyrocyte function and growth. *Endocrine Reviews* **13**, 596–611.

Zhang J., Yang J., Jang R., Zhang Y. (2015) GPCR-I-TASSER: a hybrid approach to g protein-coupled receptor structure modeling and the application to the human genome. *Structure* **23**, 1538–49.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. RMSDs of wt and domestic (Gly558Arg mutant) TSHR transmembrane domain are reported in black and red respectively.

Figure S2. Maximum helix bend angle (in degrees) calculated according to BendIx (see Methods) for TSHR TM helices 1, 3, 5 and 6 as a function of simulation time (in nanoseconds).

Figure S3. Eigenvector fluctuations along eigenvectors 1–3 are shown for RJF and GGD in black and red respectively.

Movie S1. Projection of trajectory along ED eigenvector 2 for GGD and RJF in normal and transparent modes respectively.