RESEARCH ARTICLE

Comparative analysis of human and bovine thyroglobulin structures

Han-ul Kim[1,2], Hyeongseop Jeong[3], Jeong Min Chung[4], Dooil Jeoung[1], Jaekyung Hyun[5*] and Hyun Suk Jung[1,2*]

Abstract

In biology, evolutionary conserved protein sequences show homologous physiological phenotypes in their structures and functions. If the protein has a vital function, its sequence is usually conserved across the species. However, in highly conserved protein there still remains small differences across the species. Upon protein–protein interaction (PPI), it is observed that the conserved proteins can have different binding partners that are considered to be caused by the small sequence variations in a specific domain. Thyroglobulin (TG) is the most commonly found protein in the thyroid gland of vertebrates and serves as the precursor of the thyroid hormones, tetraiodothyronine and triiodothyronine that are critical for growth, development and metabolism in vertebrates. In this study, we comparatively analyzed the sequences and structures of the highly conserved regions of TG from two different species in relation to their PPIs. In order to do so, we employed SIM for sequence alignment, STRING for PPI analysis and cryo-electron microscopy for 3D structural analysis. Our Cryo-EM model for TG of Bos taurus determined at 7.1 Å resolution fitted well with the previously published Cryo-EM model for Homo sapiens TG. By demonstrating overall structural homology between TGs from different species, we address that local amino acid sequence variation is sufficient to alter PPIs specific for the organism. We predict that our result will contribute to a deeper understanding in the evolutionary pattern applicable to many other proteins.

Keywords: Protein–protein interaction, Conserved protein, Transmission electron microscopy, Cryo-electron microscopy, Thyroglobulin

Introduction

In evolutionary biology, protein sequence conservations are commonly found in all species. Conservation proves that a primeval sequence has been preserved through natural evolution. For example, there are several conserved sequences such as the ribonucleic acid (RNA) components of ribosomes in all living organisms, the homeobox sequences in eukaryotes transfer-messenger RNA (tmRNA) in bacteria and so on (Marmur et al. 1963;

Sanger 1949). In addition, there is a linkage between conservation and length of the primary structure of protein. The molecular function of a protein predominantly determines its sequence length, and the variability of the lengths reflects the multiplicity of specialized function roles for these proteins (Knight et al. 2001). Lipman et al. have previously published the study of this hypothesis by demonstrating differences in protein length in correlation with the sequence conservation in all studied organisms. The conserved proteins are commonly longer than the less conserved proteins, and the distribution ranges of the length for the poorly conserved proteins comparatively have a smaller pool (Lipman et al. 2002). The study of sequence conservation is the basic root in the entire fields of genomics, proteomics, evolutionary biology, phylogenetics, bioinformatics and structural biology.

*Correspondence: jhyu002@pusan.ac.kr; hsjung@kangwon.ac.kr

[1] Department of Biochemistry, College of Natural Sciences, Kangwon National University, Chuncheon, Gangwon-do 24341, Republic of Korea
[5] Department of Convergence Medicine, School of Medicine, Pusan National University, Yangsan, Gyeongsangnam-do 50612, Republic of Korea
Full list of author information is available at the end of the article

Kim *et al. Journal of Analytical Science and Technology*   (2022) 13:25

Page 2 of 8

Protein–protein interaction (PPI) is one of the most significant part of biological study. PPI involves physical interactions between several protein molecules caused by biochemical events through communications that carry electrostatic forces, hydrogen bonding and the hydrophobic interaction, and it mediates a number of direct contacts between amino acid residues that occur in a specific biomolecule (De Las Rivas and Fontanillo 2010; Makino and Gojobori 2007). Thus, PPI studies of a specific protein can not only help understand its regulatory mechanism of intracellular functions, but can also provide a starting point for the predictions of other functions and related biological pathways (Ding and Kihara 2019). Most proteins hardly act alone due to the regulation mechanism. Thus, many molecular regulations in the cell are accomplished by the various proteins assembled by their PPIs. Moreover, disorders in these physiological interactions are the trigger of aggregation-related diseases, such as Creutzfeldt–Jakob and Alzheimer's disease (Li et al. 2019; Schmitz et al. 2020). Therefore, it is important to investigate the PPI as a key clue to understand the physiological mechanism of the living things.

Thyroglobulin (TG) is a 660 kDa glycoprotein that exists as homodimers. It is one of the most common protein in the thyroid gland of vertebrates and serves as the precursor of the thyroid hormones, tetraiodothyronine (T4) and triiodothyronine (T3) that are critical for growth, development and metabolism in vertebrates (Franc et al. 1990; Malthiery and Lissitzky 1987). In the thyroid gland, the formation of these hormones from TG occurs through iodination and coupling of pairs of tyrosine residues terminated through TG proteolysis (Citterio et al. 2018). TG is a most highly conserved protein in all vertebrates including the Homo sapiens, Bos taurus, Mus musculus, Canis familiaris, Sus scrofa, Danio rerio and many more (Belkadi et al. 2012; Holzer et al. 2016). Therefore, TG is mainly studied for evolutionary research due to its extremely conserved nature to understand functional and structural diversity. Decades ago, Yang et al. compared the glycosylation of TG between Homo sapiens and Bos taurus which is a main process for the function of this protein (Yang et al. 1996). In addition, Molina et al. studied the type-1 repeat from human TG, a cystein-rich module, by comparing the differences among homologous protein families (Molina et al. 1996).

Here, we comparatively analyzed the differences in the sequences and structures in the PPI action of the highly conserved TGs from Homo sapiens and Bos taurus. In order to do so, we employed the SIM alignment tool for sequence alignment (Huang et al. 1990), STRING database (Szklarczyk et al. 2019) for PPI analysis to find difference in interaction. We also performed cryo-electron microscopy (Cryo-EM) for 3D structural analysis to investigate structural variance caused by poorly conserved amino acid sequence. Our results exemplify that local amino acid sequence variations are sufficient for alteration of PPIs despite overall structural homology (Coscia et al. 2020).

Materials and methods

Sequence analysis and PPI map

Amino acids sequences of human (Homo sapiens, UniProt-P01266) and bovine (Bos taurus, UniProt-P01267) TG were adopted from UniProt database (UniProt 2021). Protein sequence alignment was performed using the alignment tool SIM (Huang 1990) and visualized by LANVIEW (Duret et al. 1996), a graphical viewer software for pairwise alignments. Statistical procedures were performed using SigmaPlot (Systat Software, San Jose, CA, US). The STRING (Szklarczyk 2019) Genes/Proteins database was used to construct the PPI networks of human and bovine TG. All the interactions between them were derived from previous information in curated databases at high level of average local clustering coefficient (human TG, 0.906; bovine TG, 0.777).

Electrostatic charge distribution analysis

For calculating electrostatic charge distribution of human and bovine TGs, a human TG atomic model was adopted from previously published paper (Coscia 2020) and an atomic model of bovine TG was constructed by SWISS-MODEL (Schwede et al. 2003), a server for template-based automated modeling of three-dimensional (3D) protein structures. The Poisson–Boltzmann equation solver program CHARMM-GUI (Jo et al. 2008) was employed to calculate the electrostatic distribution of each TG. The measured surface representation of each TGs was visualized using PyMOL (The PyMOL Molecular Graphics System, version 2.0 Schrödinger, LLC).

Transmission electron microscopy and single-particle image processing of bovine TG

Cryo-EM was performed using purified bovine TG prepared from a protein stock (T9145; Sigma-Aldrich, USA). The stock protein was dissolved in phosphate-buffered saline (10 mM phosphate buffer, 2.7 mM KCl, and 137 mM NaCl; pH 7.4) and diluted with 20 mM Tris–HCl buffer (pH 7.5) to a final concentration of 1 mg/ml. The frozen-hydrated specimen was prepared on glow-discharged Quantifoil R 1.2/1.3 holey carbon EM grids (Quantifoil, Grosslo¨ bichau, Germany) using a Vitrobot Mark IV (FEI, US; 5 s blotting time and 100% humidity at 4 °C) (Kwon et al. 2019). Automated data collection was performed using EPU software (FEI, US) by Titan Krios G2 transmission electron microscope (FEI, US) operated at 300 kV with a K2 direct electron detector (Gatan, USA)

Kim *et al. Journal of Analytical Science and Technology*   (2022) 13:25

Page 3 of 8

(instrumentation installed at Leiden University, Leiden, Netherlands). Each micrograph was recorded with a total dose of 51.02 e- Å-2 per micrographs and a defocus range from 1.5 to 3 μm at a 1.4 Å pixel size (nominal magnification of 59,000). After estimating the contrast transfer function (CTF) using Gctf (Zhang 2016) without motion correction, 1,038 micrographs with values of maximum resolution better than 5 Å resolution and defocus lower than 3 μm were selected for further processing. After performing 5 rounds of 2D classification, the best-looking 29,686 particles and 25 class averages were selected as judged by visual inspection. For reconstruction of bovine TG, initial reference model with C2 symmetry imposition was built from 3,243 particles that belong in the best 2D class averages. A total of 29, 686 particles were then selected from the best matched classes with a reference model and were subjected to 3D auto-refinement with C2 symmetry imposition. The final model was refined with a soft-edged mask and was sharpened with a -288.384 Å-2 B-factor. RELION-2.1 (Kimanius et al. 2016) package was employed to whole image processing procedures. The resolution of the final Cryo-EM reconstruction was estimated by Fourier shell correlation (FSC) between the two halves of the dataset using the FSC validation server of the Electron Microscopy Data Bank (EMDB). The local resolution was calculated with RELION-2.1. The final Cryo-EM model of bovine TG was deposited in the EMDB (EMD-30876) (Additional file 1: Table 1). UCSF Chimera (Pettersen et al. 2004) was used to superpose human TG atomic model (PDB: 6SCJ) (Coscia 2020) into bovine TG model and visual inspection. Image processing was performed using computation resources at Kangwon Center for Systems Imaging.
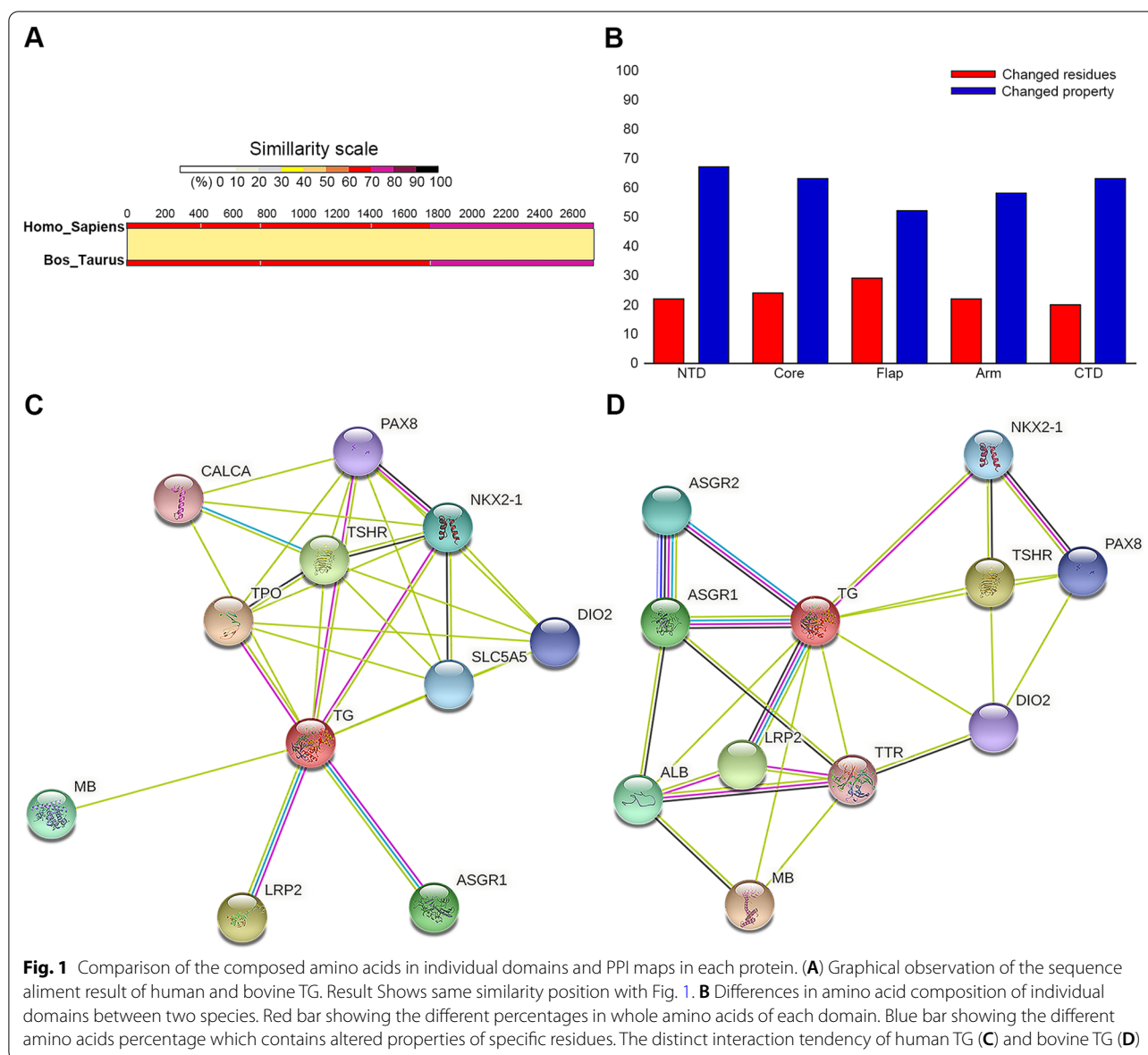
Results and discussion

Protein interaction networks of human and bovine TG

First, we conducted a sequence alignment to confirm an evolutional variance between the protein sequences of human and bovine TG (Fig. 1. A and Additional file 2: Fig. 1). As expected, following the previous research, the two proteins were observed to have high retention rates, with approximately 78% of the conserved sequences except for the cholinesterase-like (ChEL) domain of C-terminal (Fig. 1A). ChEL domain has a Cys-rich module that secretes TG in thyroid hormonogenesis in all vertebrates (Lee et al. 2008). In addition, several published papers report that hypothyroidism in human and rodents are caused by mutations in the ChEL domain (Hishinuma et al. 2000; Kim et al. 2000). Therefore, the ChEL domain is a very important part of TG and is highly conserved in all species of TG. We have identified 80% of sequence homology between humans and bovine TGs. Surprisingly, the residues that are known to be involved

in the electron acceptor/donor of the human TG during the hormonogenesis maturation have remained almost unchanged. Next, we used the sequence alignment results to understand more about the changed amino acid residues and how much they have changed in each domain (Coscia 2020). The percentage of residues that have substituted were NTD-22%, Core-24%, Flap-29%, Arm-22%, and CTD-20%, with the most changed domain being Flap and the least being CTD. Simultaneously, Among the substituted amino acids, the percentages of the changed electrostatic properties of residues were identified as NTD-67%, Core-63%, Flap-52%, Arm-58%, and CTD-63% (Fig. 1B and Additional file 3: Table 2). These changes in residues, especially in the change of electrostatic properties, are currently not understood and are expected to result from structural and functional differences for interaction with other proteins, requiring more detailed study at each domain level. Further, when we confirmed the results of the PPI network analysis, it was observed that the two proteins have a slightly different binding partner even though they have highly conserved properties (Fig. 1C and D). The PPI map of human TG shows interaction with calcitonin-related polypeptide alpha (CALCA), thyroid peroxidase (TPO) and sodium/iodide cotransporter (SLC5A5), which are not in the map of bovine TG. On the other hand, interactions with albumin (ALB), asialoglycoprotein receptor2 (ASGR2) and transthyretin (TTR) were identified in bovine TG. Following this result, the different interaction tendency between two highly conserved proteins is prospected to evolutionary distinct properties for physiological activity, and this difference is thought to be accompanied by changes in the properties of amino acid residue sequences.

The structure model of bovine TG

Next, we reconstructed a 3D model of bovine TG using Cryo-EM to demonstrate structural similarity between TGs across species despite of different PPI tendency based on sequence analysis. Our Cryo-EM map was reconstructed with 7.1 Å resolution from 29,686 particles, and we estimated the local resolution by RELION (Kimanius 2016) (Additional file 4: Fig. 2 and Additional file 1: Table 1). To compare with the previously reported 3D structure of human TG (PDB: 6SCJ) (Coscia 2020), these maps were superposed (Fig. 2). First, the overall structural difference could not be identified at given resolutions of the bovine TG (7.1 Å, blue mesh) and the human TG (3.6 Å, green) (Fig. 2C). 3D structures of the two proteins, as expected by the high sequence homology, were similar, especially in rigid regions. It is also predicted that the function of the proteins associated with the identical structural features

Fig. 1 Comparison of the composed amino acids in individual domains and PPI maps in each protein. (**A**) Graphical observation of the sequence aliment result of human and bovine TG. Result Shows same similarity position with Fig. 1. **B** Differences in amino acid composition of individual domains between two species. Red bar showing the different percentages in whole amino acids of each domain. Blue bar showing the different amino acids percentage which contains altered properties of specific residues. The distinct interaction tendency of human TG (**C**) and bovine TG (**D**)

will have the same physiological characteristics across the species (Lee et al. 2007). Therefore, it is speculated that sequence variations of amino acid residues in the same binding sites are used for interaction with different binding partners for distinct molecular functions although no significant structural differences are noticeable (Johansson-Akhe et al. 2019). In addition, we fitted the atomic model of human TG into our Cryo-EM map of bovine TG because of insufficient map resolution for direct atomic model refinement (Figs. 2D and 3). The result clearly indicated that there was no large structural difference.

The effects of amino acid substitutions on PPI

First, an electrostatic surface of two TGs was established to identify the differences, such as hydrophobicity and polarity of the amino acids, known to be involved in protein molecular interaction (Fig. 4) (Brock et al. 2007). The result demonstrates that general tendency of the surface electrostatic potential distributions is almost similar for both proteins. However, we observed small local variations. In addition, when we examined the sequence alignment result, there were a few areas where amino acid properties changed, accompanied by differences in more than 3 consecutive
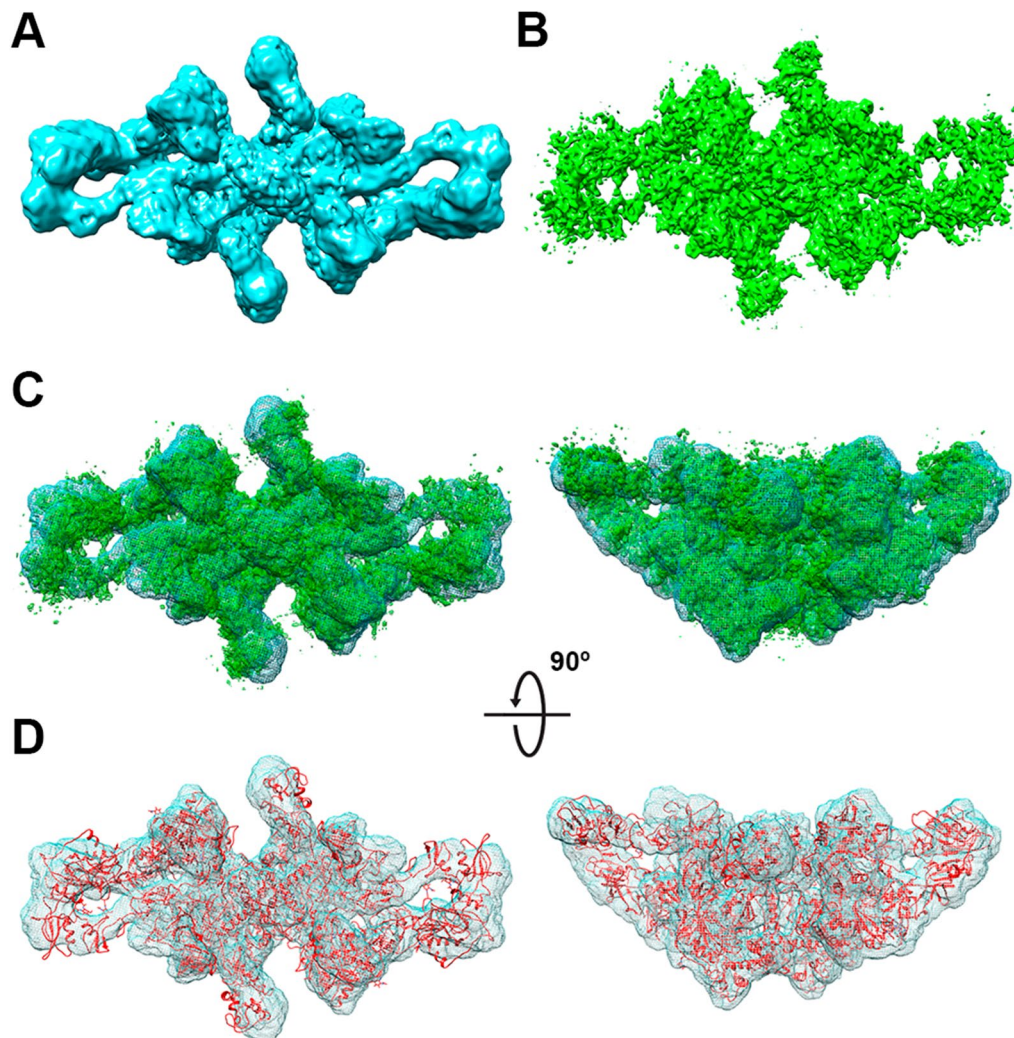
Fig. 2 Comparison of Cryo-EM maps of bovine and human TG. Refined Cryo-EM map of (**A**) bovine TG and (**B**) human TG (EMD-10141). (**C**) Overall deviations between the reconstruction map from bovine (cyan) and human (green) TG density maps (mesh representation). (**D**) The human TG atomic model (PDB: 6SCJ) fitted into the 5.1 Å bovine TG Cryo-EM reconstruction map with C2 symmetry

amino acid residues (Additional file 3: Table 2). Simultaneously, most of the changed residues were located in the loop region in each domains known as an essential role in protein–protein binding (Shehu and Kavraki 2012). Figure 5 shows local distribution of the amino acids with electrostatic residues with alternating charge properties which are predominantly located in the edge loop region in individual domains. In addition, due to the relatively low resolution of reconstruction result in this paper, we compared and verified the

3D map of bovine TG, which was recently published as high resolution, with our result and human TG map, and confirmed that there was no structural heterogeneity (Additional file 5: Fig. 4) (Kim et al. 2021). Therefore, considering the structural and physiological feature of both TGs, we propose that changes in the electrostatic properties of these small number of amino acids, especially for exposed loop regions, could affect PPI.
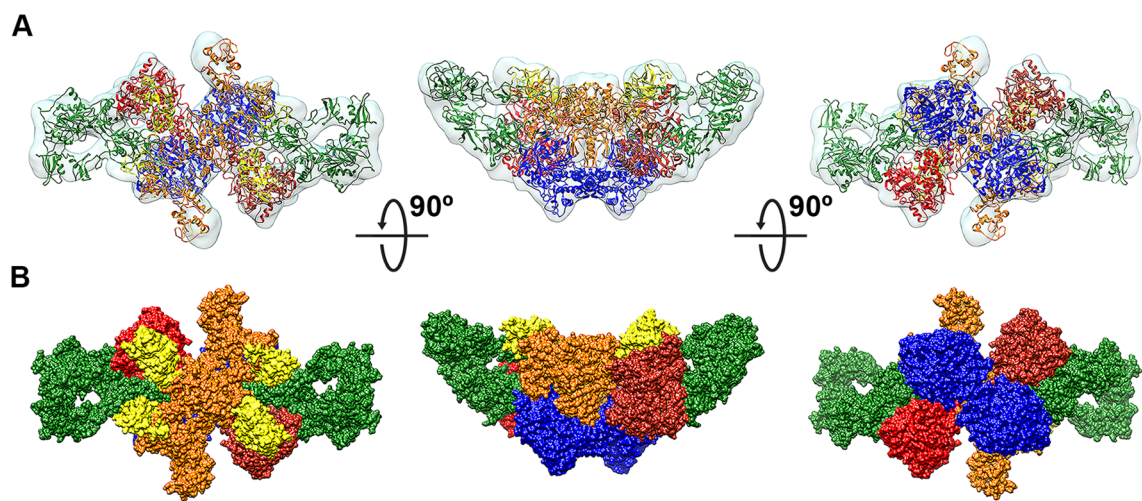
Kim *et al. Journal of Analytical Science and Technology*   (2022) 13:25

Page 6 of 8

Fig. 3 Superposition of TG atomic model from human onto the bovine Cryo-EM map. (**A**) Superposed human TG atomic model in the bovine TG EM model, and (**B**) the surface representation of the human TG (PDB: 6SCJ). TG domains are colored by NTD: red. Core: orange. Flap: yellow. Arm: green. CTD: blue
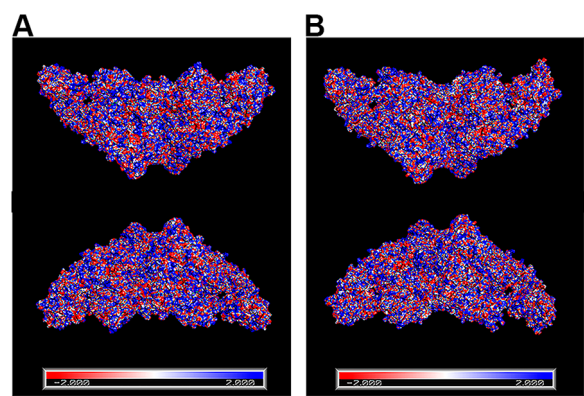

Fig. 4 Electrostatic potential surface map comparison of human and bovine TG. Calculated electrostatic potential maps from CHARMM-GUI of (**A**) human TG and (**B**) bovine TG by PyMOL. Blue color represents positive potential, white color represents neutral potential and red color represents negative potential. The potential scale used ranged from −2.000 K B T/e (red) to + 2.000 K B T/e (blue)

Conclusion

In this study, we focused on the differences between human and bovine TGs with respect to their Cryo-EM structures, electrostatic potential distribution, comparative sequence analysis and each respective PPIs. The results demonstrate the notion that despite the homology between the structures, partial changes in amino acid residues and corresponding changes in local environment can affect PPIs, ultimately leading to overall physiological regulation at molecular and cellular levels. We expect our study will not only provide a further understanding of TG, but also suggest a treatment strategy against the diseases caused by abnormal PPI activations, and also may facilitate evolutionary predictions of the cause of the PPI differences among homologous proteins.
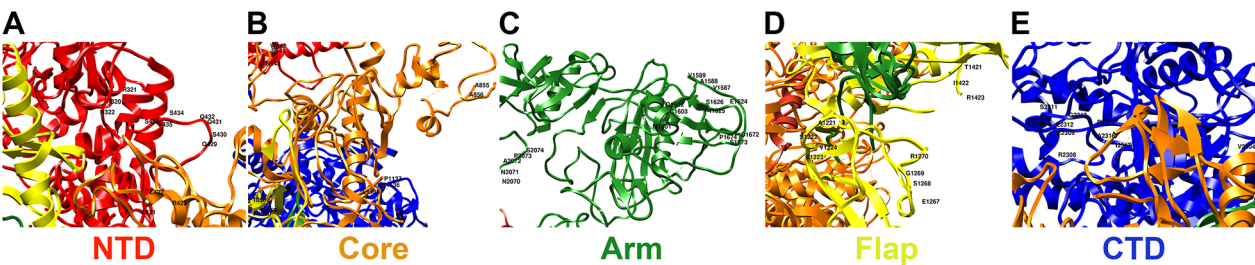

Fig. 5 Position of amino acid residues changed in human and bovine TG. (**A–E**) Amino acid residue distribution in the edge loop region of each domain containing continuously variable residues of electrostatic properties remarked in Additional file 3: Table 2, based on sequence alignment between human TG to bovine TG

Kim *et al. Journal of Analytical Science and Technology*   (2022) 13:25

Page 7 of 8

Abbreviations
RNA: Ribonucleic acid; tmRNA: Transfer-messenger RNA; PPI: Protein–protein interaction; TG: Thyroglobulin; T4: Tetraiodothyronine; T3: Triiodothyronine; Cryo-EM: Cryo-electron microscopy; 3D: Three-dimensional; CTF: Contrast transfer function; FSC: Fourier shell correlation; ChEL: Cholinesterase-like; CALCA: Calcitonin-related polypeptide alpha; TPO: Thyroid peroxidase; SLC5A5: Sodium/iodide cotransporter; ALB: Albumin; ASGR2: Asialoglycoprotein receptor 2; TTR: Transthyretin.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s40543-022-00330-9.

> **Additional file 1: Table 1.** Cryo-EM data collection and model refinement statistics.
>
> **Additional file 2: Figure 1.** Sequence alignment of TGs. The sequence alignment of TG form two vertebrate species; human and bovine. Position with >80% of conserved residues are marked with bigger letters. The plus symbols indicate no sequence similarity.
>
> **Additional file 3: Table 2.** Difference of unconserved amino acid changes between two TGs.
>
> **Additional file 4: Figure 2.** Cryo-EM data processing workflow for the bovine TG. (A) Details of cryo-EM data processing. Images show a representative micrograph (left) and 2D class averages (right). The workflow summarizes the image pre-processing (CTF estimation, particle picking), 2D classification and 3D refinement of bovine TG. The final map was refined with C2 symmetry imposition and sharpened with a B factor -288.384 Å2 (Scale bars: right, 50 nm; left, 20 nm;). (B) A montage of orthogonal views of the cryo-EM map (C) Fourier shell correlation (FSC) plots indicating resolutions of 7.1 Å for the final reconstructions according to the FSC 0.143 criterion. (D) Local resolution estimation and (E) Angular distribution plot for the final reconstruction.
>
> **Additional file 5: Figure 3.** Comparison of high-resolution cryo-EM maps of bovine (EMD-24181) and human TG (EMD-10141). Reconstructed cryo-EM map of (A; EMD-30876) bovine TG from this paper (cyan) and (B) high-resolution map (EMD-24181). Structural surface deviations of both bovine TG maps in (C) and high-resolution bovine (red) and human TG (green; (EMD-10141) maps in (D).

Acknowledgements
Not applicable

Availability of data and materials
All data generated or analyzed during this study are included in this article, and no datasets were generated or analyzed during the current study.

Declarations

Competing interests
No potential competing interest relevant to this article was reported.

Author details
[1]Department of Biochemistry, College of Natural Sciences, Kangwon National University, Chuncheon, Gangwon-do 24341, Republic of Korea. [2]Kangwon Center for Systems Imaging, Chuncheon, Gangwon-do 24341, Republic of Korea. [3]Center for Electron Microscopy Research, Korea Basic Science Institute, Ochang, Chungcheongbuk-do 28119, Republic of Korea. [4]Department of Biotechnology, The Catholic University of Korea, Bucheon-si, Gyeonggi-do 14662, Republic of Korea. [5]Department of Convergence Medicine, School of Medicine, Pusan National University, Yangsan, Gyeongsangnam-do 50612, Republic of Korea.

References
Belkadi A, Jacques C, Savagner F, Malthiery Y. Phylogenetic analysis of the human thyroglobulin regions. Thyroid Res. 2012;5:3. https://doi.org/10.1186/1756-6614-5-3.

Brock K, Talley K, Coley K, Kundrotas P, Alexov E. Optimization of electrostatic interactions in protein-protein complexes. Biophys J. 2007;93:3340–52. https://doi.org/10.1529/biophysj.107.112367.

Citterio CE, Morishita Y, Dakka N, Veluswamy B, Arvan P. Relationship between the dimerization of thyroglobulin and its ability to form triiodothyronine. J Biol Chem. 2018;293:4860–9. https://doi.org/10.1074/jbc.RA118.001786.

Coscia F, Taler-Vercic A, Chang VT, Sinn L, O'Reilly FJ, Izore T, Renko M, Berger I, Rappsilber J, Turk D, Lowe J. The structure of human thyroglobulin. Nature. 2020;578:627–30. https://doi.org/10.1038/s41586-020-1995-4.

De Las RJ, Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol. 2010;6: e1000807. https://doi.org/10.1371/journal.pcbi.1000807.

Ding Z, Kihara D. Computational identification of protein-protein interactions in model plant proteomes. Sci Rep. 2019;9:8740. https://doi.org/10.1038/s41598-019-45072-8.

Duret L, Gasteiger E, Perriere G. LALNVIEW: a graphical viewer for pairwise sequence alignments. Comput Appl Biosci. 1996;12:507–10. https://doi.org/10.1093/bioinformatics/12.6.507.

Franc JL, Venot N, Marriq C. Characterization of the two oligosaccharides present in the preferential hormonogenic domain of human thyroglobulin. Biochem Biophys Res Commun. 1990;166:937–44. https://doi.org/10.1016/0006-291x(90)90901-x.

Hishinuma A, Furudate S, Oh-Ishi M, Nagakubo N, Namatame T, Ieiri T. A novel missense mutation (G2320R) in thyroglobulin causes hypothyroidism in rdw rats. Endocrinology. 2000;141:4050–5. https://doi.org/10.1210/endo.141.11.7794.

Holzer G, Morishita Y, Fini JB, Lorin T, Gillet B, Hughes S, Tohme M, Deleage G, Demeneix B, Arvan P, Laudet V. Thyroglobulin represents a novel molecular architecture of vertebrates. J Biol Chem. 2016;291:16553–66. https://doi.org/10.1074/jbc.M116.719047.

Huang XQ, Hardison RC, Miller W. A space-efficient algorithm for local similarities. Comput Appl Biosci. 1990;6:373–81. https://doi.org/10.1093/bioinformatics/6.4.373.

Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29:1859–65. https://doi.org/10.1002/jcc.20945.

Johansson-Akhe I, Mirabello C, Wallner B. Predicting protein-peptide interaction sites using distant protein complexes as structural templates. Sci Rep. 2019;9:4267. https://doi.org/10.1038/s41598-019-38498-7.

Kim PS, Ding M, Menon S, Jung CG, Cheng JM, Miyamoto T, Li B, Furudate S, Agui T. A missense mutation G2320R in the thyroglobulin gene causes non-goitrous congenital primary hypothyroidism in the WIC-rdw rat. Mol Endocrinol. 2000;14:1944–53. https://doi.org/10.1210/mend.14.12.0571.

Kim *et al. Journal of Analytical Science and Technology*   (2022) 13:25

Page 8 of 8

Kim K, Kopylov M, Bobe D, Kelley K, Eng ET, Arvan P, Clarke OB. The structure of natively iodinated bovine thyroglobulin. Acta Crystallogr D Struct Biol. 2021;77:1451–9. https://doi.org/10.1107/S2059798321010056.

Kimanius D, Forsberg BO, Scheres SH, Lindahl E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife. 2016. https://doi.org/10.7554/eLife.18722.

Knight RD, Freeland SJ, Landweber LF. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. 2001. https://doi.org/10.1186/gb-2001-2-4-research0010.

Kwon E, Pathak D, Kim HU, Dahal P, Ha SC, Lee SS, Jeong H, Jeoung D, Chang HW, Jung HS, Kim DY. Structural Insights into stressosome assembly. IUCrJ. 2019;6:938–47. https://doi.org/10.1107/S205225251900945X.

Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol. 2007;8:995–1005. https://doi.org/10.1038/nrm2281.

Lee J, Di Jeso B, Arvan P. The cholinesterase-like domain of thyroglobulin functions as an intramolecular chaperone. J Clin Invest. 2008;118:2950–8. https://doi.org/10.1172/JCI35164.

Li M, Luo Z, Yu S, Tang Z. Proton pump inhibitor use and risk of dementia: systematic review and meta-analysis. Medicine (baltimore). 2019;98: e14422. https://doi.org/10.1097/MD.0000000000014422.

Lipman DJ, Souvorov A, Koonin EV, Panchenko AR, Tatusova TA. The relationship of protein conservation and sequence length. BMC Evol Biol. 2002;2:20. https://doi.org/10.1186/1471-2148-2-20.

Makino T, Gojobori T. Evolution of protein-protein interaction network. Genome Dyn. 2007;3:13–29. https://doi.org/10.1159/000107601.

Malthiery Y, Lissitzky S. Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. Eur J Biochem. 1987;165:491–8. https://doi.org/10.1111/j.1432-1033.1987.tb11466.x.

Marmur J, Falkow S, Mandel M. New approaches to bacterial taxonomy. Ann Rev Microbiol. 1963;17:329–72. https://doi.org/10.1146/annurev.mi.17.100163.001553.

Molina F, Bouanani M, Pau B, Granier C. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. Eur J Biochem. 1996;240:125–33. https://doi.org/10.1111/j.1432-1033.1996.0125h.x.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.

Sanger F. Species differences in insulins. Nature. 1949;164:529. https://doi.org/10.1038/164529a0.

Schmitz M, Candelise N, Kanata E, Llorens F, Thune K, Villar-Pique A, da Silva Correia SM, Dafou D, Sklaviadis T, Appelhans D, Zerr I. Validation of Poly(Propylene Imine) glycodendrimers towards their anti-prion conversion efficiency. Mol Neurobiol. 2020;57:1863–74. https://doi.org/10.1007/s12035-019-01837-w.

Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucl Acids Res. 2003;31:3381–5. https://doi.org/10.1093/nar/gkg520.

Shehu A, Kavraki LE. Modeling structures and motions of loops in protein molecules. Entropy. 2012;14:252–90.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucl Acids Res. 2019;47:D607–13. https://doi.org/10.1093/nar/gky1131.

UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucl Acids Res. 2021;49:D480–9. https://doi.org/10.1093/nar/gkaa1100.

Yang SX, Pollock HG, Rawitch AB. Glycosylation in human thyroglobulin: location of the N-linked oligosaccharide units and comparison with bovine thyroglobulin. Arch Biochem Biophys. 1996;327:61–70. https://doi.org/10.1006/abbi.1996.0093.

Zhang K. Gctf: real-time CTF determination and correction. J Struct Biol. 2016;193:1–12. https://doi.org/10.1016/j.jsb.2015.11.003.

Publisher's Note