 doi:10.20944/preprints202003.0304.v1

Article

Ab Initio Modelling the Structure of Proton-Sensing G-Protein Coupled Receptor GPR151

Wei Li*

[1] Institute of Special Environment Medicine, Nantong University, No. 9, Seyuan Road, Nantong City, Jiangsu Province, P. R. China

* Correspondence: wli148@aucklanduni.ac.nz

Abstract: Protein is the *proteios* building block of life. Evolutionarily, its sequence is not as conserved as its structure, making it more reasonable for protein structure, instead of protein sequence, to be the descriptor of protein function. Yet, in the National Center for Biotechnology Information (NCBI) database, the number of experimentally identified protein sequences is in great excess of that of experimentally determined protein structures inside the almost-half-a-century old Protein Data Bank (PDB). For instance, GPR151 is an proton-sensing G-protein coupled receptor (GPCR) originally identified as homologous to galanin receptors. As of March 19, 2020, GPR151's structure has not been experimentally determined and deposited in PDB yet. Thus, an *ab initio* modelling approach was employed here to build a three-dimensional structure of GPR151. Overall, the *ab initio* GPR151 model presented herein constitutes the first structural hypothesis of GPR151 to be experimentally tested in future with previously published, currently ongoing and future GPR151 studies.

Keywords: *Ab Initio* Modelling; Three-Dimensional Structure; Proton-Sensing G-Protein Coupled Receptor (GPCR); GPR151

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2020 doi:10.20944/preprints202003.0304.v1

1. Introduction

It has been almost half a century since the launch of Protein Data Bank (PDB) in 1971 [1]. Biophysical tools such as X-ray crystallography, NMR spectroscopy and Cryo-electron microscopy have contributed enormously to the continued development of PDB [2], with which a variety of computational tools have been developed for biomolecular structural modelling, classification and feature extraction [3–12] and functional prediction. For instance, *ab initio* protein structural modelling [13,14] is an energy function-guided method to predict protein structure in the absence of experimentally solved structure of a similar/homologous protein. Also termed as *de-novo* modelling, physics-based modelling or free modelling, the *ab initio* approach is often preferred for structure prediction when there is no or very low amount of similarity for the query protein sequence.

2. Motivation

Protein is the *proteios* building block of life. Evolutionarily, its sequence is not as conserved as its structure, making it more reasonable for protein structure, instead of protein sequence, to be the descriptor of protein function. Yet, in the NCBI database, the number of experimentally identified protein sequences is in great excess of that of experimentally determined protein structures inside the almost-half-a-century old PDB [1]. For instance, GPR151 is an proton-sensing G-protein coupled receptor (GPCR) originally identified as homologous to galanin receptors [15–30].

As of March 19, 2020, GPR151's structure has not been experimentally determined and deposited in PDB yet. Furthermore, with a structural search by the SwissModel server (https://swissmodel.expasy.org/interactive) [31], it turned out that there is no experimentally determined structure of GPR151-homologous protein in PDB [1], and that the highest similarity possible for GPR151 is only 17.33% for the chain A of Neurotensin receptor type 1 (PDB ID: 4XES) as of March 19, 2020. As a result, homology structural modelling is not an option feasible here for GPR151. Therefore, this article employs an *ab initio* modelling approach to build a three-dimensional structure of GPR151.

3. Materials and Methods

With a set of online resources (**https://www.ncbi.nlm.nih.gov/protein/EAW61839.1** and **https://www.uniprot.org/uniprot/Q8TDV0**), the amino acid sequence of GPR151 was retrieved and listed below,

MLAAAFADSNSSSMNVSFAHLHFAGGYLPSDSQDWRTIIPALLVAVCLVGFVGNLCVIGILLHN AWKGKPSMIHSLILNLSLADLSLLLFSAPIRATAYSKSVWDLGWFVCKSSDWFIHTCMAAKSLTIVVV AKVCFMYASDPAKQVSIHNYTIWSVLVAIWTVASLLPLPEWFFSTIRHHEGVEMCLVDVPAVAEEF MSMFGKLYPLLAFGLPLFFASFYFWRAYDQCKKRGTKTQNLRNQIRSKQVTVMLLSIAIISALLWLPE WVAWLWVWHLKAAGPAPPQGFIALSQVLMFSISSANPLIFLVMSEEFREGLKGVWKWMITKKPPTV SESQETPAGNSEGLPDKVPSPESPASIPEKEKPSSPSSGKGKTEKAEIPILPDVEQFWHERDTVPSVQD NDPIPWEHEDQETGEGVK

The *ab initio* construction of GPR151's structure started from its sequence consisting of 419 amino acid residues as above, and employed the Quark *ab initio* modelling server (https://zhanglab.ccmb.med.umich.edu/QUARK/) [13,14], which was ranked as the No. 1 server in Free-modelling (FM) in CASP9 and CASP10 experiments, making it suitable for proteins that do not have homologous templates (like GPR151) in the PDB database [1].

Given that Quark does not accept job submission with protein sequence longer than 200 amino acid residues, the GPR151 sequence is thus manually separated into three fragments for three *ab initio* structural modelling processes.

1. *MLAAAFADSNSSSMNVSFAHLHFAGGYLPSDSQDWRTIIPALLVAVCLVGFVGNLCVIGILLH NAWKGKPSMIHSLILNLSLADLSLLLFSAPIRATAYSKSVWDLGWFVCKSSDWFIHTCMAAKSL TIVVVAKVCFMYASDPAKQVSIHNYTIWSVLVAIWTV*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2020 doi:10.20944/preprints202003.0304.v1

2. *SDPAKQVSIHNYTIWSVLVAIWTVASLLPLPEWFFSTIRHHEGVEMCLVDVPAVAEE FMSMFGKLYPLLAFGLPLFFASFYFWRAYDQCKKRGTKTQNLRNQIRSKQVTVM LLSIAIISALLWLPEWVAWLWVVWHLKAAGPAPPQGFIALSQVLMFSISSANPLIFL VMSEEFREGLKGVWKWMITKKPPTVSESQET*
3. *MSEEFREGLKGVWKWMITKKPPTVSESQETPAGNSEGLPDKVPSPESPASIPEKEKPSSPSSGK GKTEKAEIPILPDVEQFWHERDTVPSVQDNDPIPWEHEDQETGEGVK*

Specifically, the *ab initio* construction of GPR151's structure consists of seven steps as below,

1. The *ab initio* construction of GPR151's first fragment was performed with the QUARK server with an output PDB file (supplementary file **model1.pdb**).
2. The *ab initio* construction of GPR151's second fragment was performed with the QUARK server with an output PDB file (supplementary file **model2.pdb**).
3. The *ab initio* construction of GPR151's third fragment was performed with the QUARK server with an output PDB file (supplementary file **model3.pdb**).
4. An in-house python script (supplementary file **k1.py**) was used to dock **model2.pdb** (red fragment) to **model1.pdb** (red fragment), yielding an *ab initio* modelled structural fragment (from Met1 to Thr338) of GPR151 (supplementary file **k1.pdb**).
5. **k1.pdb** was subsequently subject to an energy minimization process on the ModRefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) [32], yielding an *ab initio* modelled energy-minimized structural fragment (from Met1 to Thr338) of GPR151 (supplementary file **k1refine.pdb**).
6. An in-house python script (supplementary file **k2.py**) was used to dock **k1refine.pdb** (blue fragment) to **model3.pdb** (blue fragment), yielding an *ab initio* modelled full-length structure of GPR151 (supplementary file **k2.pdb**).
7. **k2.pdb** was subsequently subject to an energy minimization process on the ModRefiner server, yielding an *ab initio* modelled energy-minimized full-length structure of GPR151 (supplementary file **k2refine.pdb**)

4. Result and Conclusion

With the *ab initio* structural modelling steps described above, Figure 1 presents an overall view of the three-dimensional scaffold of GPR151.

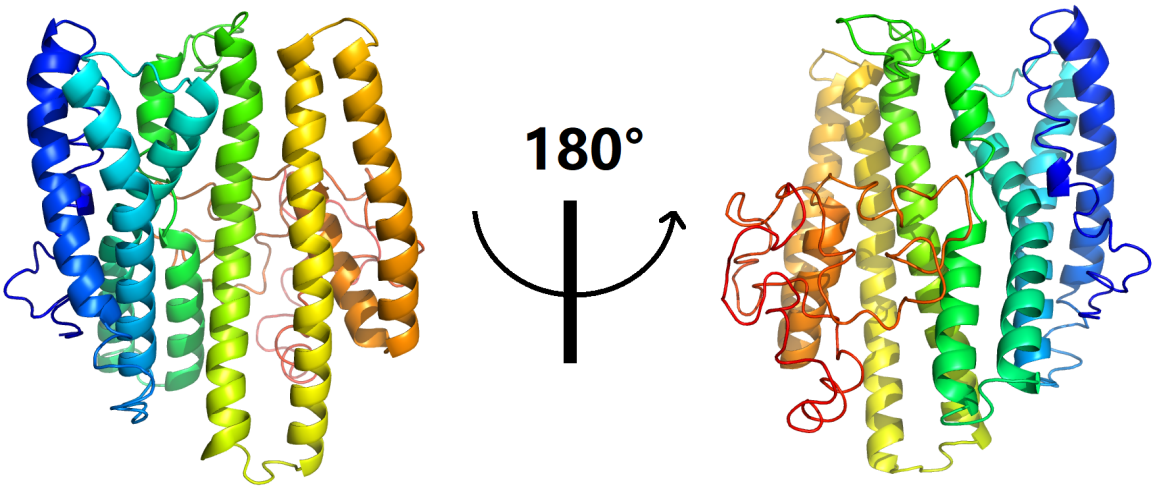

Figure 1. An *ab initio* structure of GPR151. In this figure, GPR151 is shown as rainbow-coloured cartoons, with its N- and C-termini coloured blue and red, respectively. This figure is prepared using PyMol [33] with supplementary file **k2refine.pdb** as an input.

The *ab initio* modelled atomic coordinates of GPR151 are included in the supplementary file **k2refine.pdb**, which does not include side chain hydrogen atomic coordinates for GPR151 residues.

With Chimera [34], nonetheless, all side chain hydrogen atoms can be placed at their respective spatial locations where the GPR151 structural model corresponds to the global free energy minimum under a given set of conditions [35,36], such that no atomic detail is missing and that the *ab initio* modelled GPR151 structure is intact, i.e., experimentally uncharted territory (EUT)-less [3].

Overall, the *ab initio* GPR151 model here constitutes the first structural hypothesis of GPR151, which remains to be tested in future with all experimental GPR151 studies, previously published, currently ongoing and future ones, too.

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, W.L.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, W.L.; data duration, W.L.; writing–original draft preparation, W.L.; writing–review and editing, W.L.; visualization, W.L.; supervision, W.L.; project administration, W.L.; funding acquisition, not applicable.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

1. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
2. Steven, A.C.; Baumeister, W. The future is hybrid. *Journal of Structural Biology* **2008**, *163*, 186–195.
3. Li, W. Visualising the Experimentally Uncharted Territories of Membrane Protein Structures inside Protein Data Bank **2020**.
4. Li, W. A Local Spherical Coordinate System Approach to Protein 3D Structure Description **2020**.
5. Li, W. Inter-Molecular Electrostatic Interactions Stabilizing the Structure of the PD-1/PD-L1 Axis: A Structural Evolutionary Perspective **2020**.
6. Li, W. Calcium Channel Trafficking Blocker Gabapentin Bound to the α-2-δ-1 Subunit of Voltage-Gated Calcium Channel: A Computational Structural Investigation **2020**.
7. Li, W. Structural Identification of the Electrostatic Hot Spots for Severe Acute Respiratory Syndrome Coronavirus Spike Protein to Be Complexed with Its Receptor ACE2 and Its Neutralizing Antibodies **2020**.
8. Li, W. Two Achilles' Heels of the Ebolavirus Glycoprotein? **2020**.
9. Li, W. Extracting the Interfacial Electrostatic Features from Experimentally Determined Antigen and/or Antibody-Related Structures inside Protein Data Bank for Machine Learning-Based Antibody Design **2020**.
10. Li, W.; Shi, G. How $Ca_V1.2$-bound verapamil blocks Ca^{2+} influx into cardiomyocyte: Atomic level views. *Pharmacological Research* **2019**, *139*, 153–157.
11. Li, W. Structural and Functional Consequences of the SMA-Linked Missense Mutations of the Survival Motor Neuron Protein: A Brief Update. In *Novel Aspects on Motor Neuron Disease [Working Title]*; IntechOpen, 2019.
12. Li, W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLOS ONE* **2017**, *12*, e0178519.
13. Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and Bioinformatics* **2012**, pp. n/a–n/a.
14. Xu, D.; Zhang, Y. Toward optimal fragment generations for ab initio protein structure assembly. *Proteins: Structure, Function, and Bioinformatics* **2012**, *81*, 229–239.
15. Ignatov, A.; Hermans-Borgmeyer, I.; Schaller, H. Cloning and characterization of a novel G-protein-coupled receptor with homology to galanin receptors. *Neuropharmacology* **2004**, *46*, 1114–1120.
16. Mashiko, M.; Kurosawa, A.; Tani, Y.; Tsuji, T.; Takeda, S. GPR31 and GPR151 are activated under acidic conditions. *The Journal of Biochemistry* **2019**, *166*, 317–322.
17. Broms, J.; Grahm, M.; Haugegaard, L.; Blom, T.; Meletis, K.; Tingström, A. Monosynaptic retrograde tracing of neurons expressing the G-protein coupled receptor GPR151 in the mouse brain. *Journal of Comparative Neurology* **2017**, *525*, 3227–3250.
18. Holmes, F.E.; Kerr, N.; Chen, Y.J.; Vanderplank, P.; McArdle, C.A.; Wynick, D. Targeted disruption of the orphan receptor GPR151 does not alter pain-related behaviour despite a strong induction in dorsal root ganglion expression in a model of neuropathic pain. *Molecular and Cellular Neuroscience* **2017**, *78*, 35–40.

19. Jiang, B.C.; Zhang, W.W.; Yang, T.; Guo, C.Y.; Cao, D.L.; Zhang, Z.J.; Gao, Y.J. Demethylation of G-Protein-Coupled Receptor 151 Promoter Facilitates the Binding of Krüppel-Like Factor 5 and Enhances Neuropathic Pain after Nerve Injury in Mice. *The Journal of Neuroscience* **2018**, *38*, 10535–10551.
20. Emdin, C.A.; Khera, A.V.; Chaffin, M.; Klarin, D.; Natarajan, P.; Aragam, K.; Haas, M.; Bick, A.; Zekavat, S.M.; Nomura, A.; Ardissino, D.; Wilson, J.G.; Schunkert, H.; McPherson, R.; Watkins, H.; Elosua, R.; Bown, M.J.; Samani, N.J.; Baber, U.; Erdmann, J.; Gupta, N.; Danesh, J.; Chasman, D.; Ridker, P.; Denny, J.; Bastarache, L.; Lichtman, J.H.; D'Onofrio, G.; Mattera, J.; Spertus, J.A.; Sheu, W.H.H.; Taylor, K.D.; Psaty, B.M.; Rich, S.S.; Post, W.; Rotter, J.I.; Chen, Y.D.I.; Krumholz, H.; Saleheen, D.; Gabriel, S.; Kathiresan, S. Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nature Communications* **2018**, *9*.
21. Broms, J.; Antolin-Fontes, B.; Tingström, A.; Ibañez-Tallon, I. Conserved expression of the GPR151 receptor in habenular axonal projections of vertebrates. *Journal of Comparative Neurology* **2014**, *523*, 359–380.
22. Tanigawa, Y.; Li, J.; Justesen, J.M.; Horn, H.; Aguirre, M.; DeBoever, C.; Chang, C.; Narasimhan, B.; Lage, K.; Hastie, T.; Park, C.Y.; Bejerano, G.; Ingelsson, E.; Rivas, M.A. Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nature Communications* **2019**, *10*.
23. Foll, B.L.; French, L. Transcriptomic Characterization of the Human Habenula Highlights Drug Metabolism and the Neuroimmune System. *Frontiers in Neuroscience* **2018**, *12*.
24. Antolin-Fontes, B.; Li, K.; Ables, J.L.; Riad, M.H.; Görlich, A.; Williams, M.; Wang, C.; Lipford, S.M.; Dao, M.; Liu, J.; Molina, H.; Heintz, N.; Kenny, P.J.; Ibañez-Tallon, I. The habenular G-protein–coupled receptor 151 regulates synaptic plasticity and nicotine intake. *Proceedings of the National Academy of Sciences* **2020**, *117*, 5502–5509.
25. Kakarala, K.K.; Jamil, K. Sequence-structure based phylogeny of GPCR Class A Rhodopsin receptors. *Molecular Phylogenetics and Evolution* **2014**, *74*, 66–96.
26. Wagner, F.; Bernard, R.; Derst, C.; French, L.; Veh, R.W. Microarray analysis of transcripts with elevated expressions in the rat medial or lateral habenula suggest fast GABAergic excitation in the medial habenula and habenular involvement in the regulation of feeding and energy balance. *Brain Structure and Function* **2016**, *221*, 4663–4689.
27. Tang, S.; Jing, H.; Huang, Z.; Huang, T.; Lin, S.; Liao, M.; Zhou, J. Identification of key candidate genes in neuropathic pain by integrated bioinformatic analysis. *Journal of Cellular Biochemistry* **2019**, *121*, 1635–1648.
28. Liu, Z.; Xu, Y.; Wu, L.; Zhang, S. Evolution of Galanin Receptor Genes: Insights from the Deuterostome Genomes. *Journal of Biomolecular Structure and Dynamics* **2010**, *28*, 97–106.
29. Li, H.L.; Lee, J.R.; Hahn, M.J.; Yang, J.M.; Meng, F.G.; Wu, J.W.; Park, Y.D. The omics based study for the role of superoxide dismutase 2 (SOD2) in keratinocytes: RNA sequencing, antibody-chip array and bioinformatics approaches. *Journal of Biomolecular Structure and Dynamics* **2019**, pp. 1–14.
30. Kaelberer, M.M.; Caceres, A.I.; Jordt, S.E. Activation of a nerve injury transcriptional signature in airway-innervating sensory neurons after Lipopolysaccharide induced lung inflammation. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2020**.
31. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.
32. Xu, D.; Zhang, Y. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophysical Journal* **2011**, *101*, 2525–2534.
33. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **2002**, *40*, 82–92.
34. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.
35. Anfinsen, C.B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.
36. Hirata, F.; Sugita, M.; Yoshida, M.; Akasaka, K. Perspective: Structural fluctuation of protein and Anfinsen's thermodynamic hypothesis. *The Journal of Chemical Physics* **2018**, *148*, 020901.