

Response to “Addressing flaws in the Seq2Topt dataset for the prediction of enzyme optimal temperature”

Sizhe Qiu ^{1,2,*} and Aidong Yang ²

¹State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai, 200237, P. R. China

²Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom

*Corresponding author. State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, P. R. China. E-mail: siq001@outlook.com

Response

To the Editor,

We would like to thank the rigorous dataset inspection work conducted by the author of “Addressing flaws in the Seq2Topt dataset for the prediction of enzyme optimal temperature”. The dataset of enzyme optimal temperature (T_{opt}) used to develop Seq2Topt [1] was obtained from <https://github.com/jafetgado/tomer>, which was curated by Li *et al.* 2019 [2] and Gado *et al.* 2020 [3] from the BRENDA database [4]. In “Addressing flaws in the Seq2Topt dataset for the prediction of enzyme optimal temperature,” the author stated that the T_{opt} values of F9FU71, A9U908, J7HET3, P94368, Q5YXQ1, Q83V33, Q50539, Q50538, Q2WCS9 are erroneously annotated in the dataset used by Seq2Topt. Here, we presented all references associated with these data entries in the BRENDA database in Table 1, and the T_{opt} values of F9FU71, P94368, Q5YXQ1, Q83V33, Q50539, and Q50538 were not found in relevant references [5–9]. Dhar *et al.* 2013 [10] stated that the T_{opt} values of A9U908 (sodA) and J7HET3 (sodB) were 4°C, but we agree with the critique that no enzyme activity measurements at lower temperatures were conducted to confirm that 4°C was the T_{opt} value of A9U908 and J7HET3. For Q2WCS9, De Angelis *et al.* 2010 [11] reported that 13°C was the measured T_{opt} value, whereas the critique letter did not provide a source for the claimed measured T_{opt} value of 37°C.

Overall, we concur with the critique that the dataset used to develop enzyme T_{opt} predictive models (and machine learning models for other protein properties) should be carefully reviewed and revised to ensure validity, rather than simply extracting data from established databases such as BRENDA. While such curation may be costly, it remains necessary until a high level of validity is automatically guaranteed by these biochemical databases. Accordingly, future updates of Seq2Topt will be based on a more rigorously vetted dataset.

Table 1 Selected nine entries from Seq2Topt dataset.

UniProt accession number	T_{opt} in Seq2Topt dataset (°C)	Reference in BRENDA database	T_{opt} in literature (°C)
F9FU71	0	[5]	Not found
A9U908	4	[10]	4
J7HET3	4	[10]	4
P94368	5	[6]	Not found
Q5YXQ1	7.5	[7]	Not found
Q83V33	10	[8]	Not found
Q50539	13	[9]	Not found
Q50538	13	[9]	Not found
Q2WCS9	13	[11]	13

Conflict of interest

The author declares that there is no conflict of interest.

Funding

None declared.

Data availability

No new data was generated in this study. This response letter refers to the data provided in the Seq2Topt study [1].

References

1. Qiu S, Hu B, Zhao J *et al.* Seq2Topt: a sequence-based deep learning predictor of enzyme optimal temperature. *Brief Bioinform* 2025;**26**:bbaf114. <https://doi.org/10.1093/bib/bbaf114>

Received: May 8, 2026. Accepted: May 11, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

2. Li G, Rabe KS, Nielsen J *et al.* Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS Synth Biol* 2019;**8**:1411–20. <https://doi.org/10.1021/acssynbio.9b00099>
3. Gado JE, Beckham GT, Payne CM. Improving enzyme optimum temperature prediction with resampling strategies and ensemble learning. *J Chem Inf Model* 2020;**60**:4098–107. <https://doi.org/10.1021/acs.jcim.0c00489>
4. Jeske L, Placzek S, Schomburg I *et al.* BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 2019;**47**:D542–9. <https://doi.org/10.1093/nar/gky1048>
5. Sooman L, Oliw EH. Discovery of a novel linoleate dioxygenase of *Fusarium oxysporum* and linoleate diol synthase of *Colletotrichum graminicola*. *Lipids* 2015;**50**:1243–52. <https://doi.org/10.1007/s11745-015-4078-9>
6. Shumilin IA, Cymborowski M, Chertihin O *et al.* Identification of unknown protein function using metabolite cocktail screening. *Structure* 2012;**20**:1715–25. <https://doi.org/10.1016/j.str.2012.07.016>
7. Fisch F, Fleites CM, Delenne M *et al.* A covalent succinylcysteine-like intermediate in the enzyme-catalyzed transformation of maleate to fumarate by maleate isomerase. *J Am Chem Soc* 2010;**132**:11455–7. <https://doi.org/10.1021/ja1053576>
8. Huo L, Davis I, Liu F *et al.* Crystallographic and spectroscopic snapshots reveal a dehydrogenase in action. *Nat Commun* 2015;**6**:5935. <https://doi.org/10.1038/ncomms6935>
9. Maupin-Furlow J, Ferry JG. Characterization of the cdhD and cdhE genes encoding subunits of the corrinoid/iron-sulfur enzyme of the CO dehydrogenase complex from *Methanosarcina thermophila*. *J Bacteriol* 1996;**178**:340–6. <https://doi.org/10.1128/jb.178.2.340-346.1996>
10. Dhar MS, Gupta V, Virdi JS. Detection, distribution and characterization of novel superoxide dismutases from *Yersinia enterocolitica* Biovar 1A. *PLoS One* 2013;**8**:e63919. <https://doi.org/10.1371/journal.pone.0063919>
11. De Angelis M, Calasso M, Di Cagno R *et al.* NADP-glutamate dehydrogenase activity in nonstarter lactic acid bacteria: effects of temperature, pH and NaCl on enzyme activity and expression. *J Appl Microbiol* 2010;**109**:1763–74. <https://doi.org/10.1111/j.1365-2672.2010.04804.x>