Briefings in Bioinformatics, 2026, 27, bbag308
https://doi.org/10.1093/bib/bbag308
Published: 14 June 2026
Letter to Editor

OXFORD

Letter to the editor: methodological considerations in the benchmarking of AI-based protein–aptamer complex prediction

Kazim Okan Dolu*

Department of Pediatric Allergy and Immunology, İstinye University, Liv Bahçeşehir Hospital, Beşyol Mah., İnönü Cad. No. 2, Küçükçekmece, 34295, Istanbul, Turkey

*Corresponding author. Department of Pediatric Allergy and Immunology, İstinye University, Liv Bahçeşehir Hospital, Beşyol Mah., İnönü Cad. No. 2, Küçükçekmece, 34295, Istanbul, Turkey. E-mail: okandolu@gmail.com

I read with interest the article by Zhao et al. on AI-based protein–aptamer complex prediction [1]. The multi-metric design is sound and the inclusion of molecular dynamics simulations adds practical value. However, I have concerns about four methodological points that affect how the results should be interpreted.

First, the Methods section contains an internal inconsistency regarding model versions. All models are stated to have been accessed in May 2025, yet Boltz-2 is listed as v2.2.0, released in July 2025 [1]. A July release cannot have been accessed in May. The authors should clarify the actual version, access date, and commit identifier used. This matters for reproducibility.

Second, the term "independent benchmark" should be used with caution. Six of the 11 benchmark complexes fall within the Boltz-2 training set [1]. The authors provide a seen/unseen split, which is appropriate. However, describing the overall benchmark as independent without qualification may lead readers to overestimate the neutrality of cross-model comparisons.

Third, the secondary-structure input used for the 3dRNA/DNA evaluation requires clarification. Nine of the 11 aptamers are DNA molecules (Table 2 in Zhao et al. [1]), yet their secondary structures were generated using RNAFold, an RNA-oriented tool [1]. Because 3dRNA/DNA relies on secondary-structure-guided template assembly, this choice may introduce systematic bias in the majority of cases. This issue deserves explicit clarification and, ideally, a sensitivity analysis. Given that 3dRNA/DNA is the sole template-based method in the comparison, DNA-compatible folding approaches using appropriate DNA parameters would have been a more suitable choice.

Fourth, the binding free energy terminology needs to be more precise. Figure 1 labels the x-axis as "Ground Truth Binding Free Energy" [1]. These values are not experimental binding affinities measured by isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR). They are Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) estimates derived from molecular dynamics (MD) trajectories of experimental structures, without entropy correction [1]. Describing computationally derived quantities as ground truth may overstate the energetic conclusions.

I also have a concern regarding the negative controls. The mismatched pairs were constructed carefully, but their nonbinding status was not experimentally confirmed [1]. In protein–nucleic acid systems, nonspecific electrostatic association is common. Describing these pairs as "genuine nonbinding relationships" is stronger than the evidence supports.

Data availability

No new data were generated or analyzed in this study.

Funding

None declared.

Reference

1. Zhao J, Tram K, Yan H *et al.* Comprehensive evaluation of artificial intelligence-empowered approaches for protein–aptamer complex prediction. *Brief Bioinform* 2026;**27**:bbag206. https://doi.org/10.1093/bib/bbag206

Received: May 4, 2026. **Accepted:** May 18, 2026