

ABSTRACT

BENCHMARKING ALPHAFOLD3 FOR PROTEIN SECONDARY STRUCTURE PREDICTION

By

I Gusti Agung Premananda

21/473829/PA/20432

Protein Secondary Structure Prediction (PSSP) is a crucial component of computational biology, yet its accuracy remains a challenge, especially for non-canonical structures. This study exploratively evaluates AlphaFold3—a state-of-the-art model for tertiary structure prediction—as a PSSP tool by converting its 3D outputs using the DSSP algorithm. The evaluation was performed on the CB513, TS115, and CASP10 benchmark datasets using precision, recall, F1-score, and Q3/Q8 accuracy metrics. The results show that AlphaFold3 establishes a new, very high-performance baseline, consistently excelling in Q3 accuracy and achieving superior F1-scores for major structure classes such as the α -helix (H) and β -strand (E). Significantly, the model demonstrates a unique generalization capability by being the only one able to detect the highly rare π -helix (I) class with a meaningful recall. However, the analysis also identifies key limitations, namely a tendency for over-prediction in minor classes, which results in low precision, and difficulty in distinguishing flexible structures like turns (T) and bends (S). These findings confirm that AlphaFold3 can serve as a very powerful PSSP tool, particularly for common structures.

Keywords: AlphaFold3, Protein Secondary Structure Prediction (PSSP), Deep Learning, Q8 Accuracy, DSSP, Bioinformatics

ABSTRAK

EVALUASI PERFORMA ALPHAFOLD3 UNTUK PREDIKSI STRUKTUR SEKUNDER PROTEIN

Oleh

I Gusti Agung Premananda

21/473829/PA/20432

Prediksi Struktur Sekunder Protein (PSSP) adalah komponen krusial dalam biologi komputasional, namun akurasinya masih menjadi tantangan, terutama untuk struktur non-kanonik. Penelitian ini secara eksploratif mengevaluasi AlphaFold3—sebuah model *state-of-the-art* untuk prediksi struktur tersier—sebagai alat PSSP dengan mengonversi output 3D-nya menggunakan algoritma DSSP. Evaluasi dilakukan pada dataset benchmark CB513, TS115, dan CASP10 menggunakan metrik *precision*, *recall*, *F1-score*, serta akurasi Q3 dan Q8. Hasil menunjukkan bahwa AlphaFold3 menetapkan *baseline* performa baru yang sangat tinggi, secara konsisten unggul dalam akurasi Q3 dan mencapai *F1-score* superior untuk kelas-kelas struktur mayor seperti α -helix (H) dan β -strand (E). Secara signifikan, model ini menunjukkan kemampuan generalisasi yang unik dengan menjadi satu-satunya yang mampu mendeteksi kelas π -helix (I) yang sangat langka dengan *recall* yang berarti. Namun, analisis juga mengidentifikasi keterbatasan utama, yaitu kecenderungan *over-prediction* pada kelas minor yang menghasilkan presisi rendah, serta kesulitan dalam membedakan struktur fleksibel seperti *turn* (T) dan *bend* (S). Temuan ini mengukuhkan bahwa AlphaFold3 dapat berfungsi sebagai alat PSSP yang sangat kuat, terutama untuk struktur umum.

Kata Kunci: AlphaFold3, Prediksi Struktur Sekunder Protein (PSSP), *Deep Learning*, Akurasi Q8, DSSP, Bioinformatika