

XÂY DỰNG MÔ HÌNH ĐÁNH GIÁ LƯỢNG PROTEIN DỰA TRÊN ĐẶC TRƯNG TRÌNH TỰ mRNA Ở *Escherichia coli* K-12 MG 1655

Trương Hà Minh Nhật^{1,2}, *Lê Tấn Thuận*^{1,2}, *Võ Trí Nam*^{1,2}, *Nguyễn Đức Hoàng*^{1,2*}

¹Trường Đại học Khoa học Tự Nhiên, ĐHQG-HCM

²Đại học Quốc gia Thành phố Hồ Chí Minh

thmnhat25496@gmail.com, tanthuan1718314.khtn.hcmus@gmail.com,
vtnam@hcmus.edu.vn, ndhoang@hcmus.edu.vn

Việc hiểu được mối tương quan giữa đặc trưng trình tự mRNA và mức độ biểu hiện protein giúp hỗ trợ các nghiên cứu sản xuất protein tái tổ hợp. Hiện nay, mô hình Transim của Tuller và cộng sự (2017) đã tính toán tốc độ khởi đầu và tốc độ kéo dài dịch mã, từ đó dự đoán được tốc độ dịch mã dựa vào trình tự mRNA. Tuy nhiên, mô hình này chưa xét đến tác động phân hủy protein lên lượng protein đầu ra. Vì vậy, nghiên cứu tiến hành khảo sát độ bất ổn định protein *in vitro* dựa trên trình tự protein, kết hợp với kết quả từ mô hình Transim để dự đoán lượng protein tạo thành. Kết quả kiểm tra trên chủng chủ *E. coli* K-12 MG1655 cho thấy mô hình mới cho mức độ tương quan Spearman giữa kết quả đầu ra của mô hình với lượng protein thực tế tăng từ 0,37 lên 0,56 so với mô hình Transim. Như vậy, nghiên cứu cung cấp một mô hình động lực dịch mã hoàn chỉnh hơn, kết hợp giữa tốc độ khởi đầu, tốc độ kéo dài và độ bền của protein.

Từ khóa: độ bền protein, *E. coli*, mRNA, tốc độ dịch mã, Transim

CONSTRUCTING A MODEL EVALUATING PROTEIN LEVELS BASED ON mRNA SEQUENCE FEATURES IN *Escherichia coli* K-12 MG1655

Truong Ha Minh Nhat^{1,2}, *Le Tan Thuan*^{1,2}, *Nam Vo*^{1,2}, *Nguyen Duc Hoang*^{1,2*}

¹University of Science, VNU-HCM

²Vietnam National University HCMC

thmnhat25496@gmail.com, tanthuan1718314.khtn.hcmus@gmail.com,
vtnam@hcmus.edu.vn, ndhoang@hcmus.edu.vn

Understanding the relationship between mRNA sequence features and protein expression levels would support researches about the production of recombinant proteins. In recent times, the Transim model developed by Tuller *et al.* (2017) was able to calculate initiation rate and elongation rate based on an input mRNA sequence. These were then used to infer translation rate. The model, however, have not considered the effect of protein degradation rate on the protein output. This research analyzed protein instability indexes *in vitro* based on protein sequences, and consequently combined the indexes with the outputs from Transim

model to predict protein levels. The results on *E. coli* K-12 MG1655 showed that the new model was able to improve the Spearman correlation between translation rate and protein levels up to 0.56, compared to 0.37 from the original Transim. Therefore, this research has provided a more completed translational model, combining initiation rate, elongation rate and protein degradation rate.

Keywords: *E. coli*, mRNA, protein instability, Transim, translation rate