FROM SEQUENCE TO STRENGTH: PREDICTION AND DESIGN OF INTRINSIC TRANSCRIPTION TERMINATORS

Guilherme E. Kundlatsch, Almiro P. S. Neto, Gabriela B. de Paiva & Danielle B. Pedrolli
Department of Bioprocess Engineering and Biotechnology
School of Pharmaceutical Sciences
Universidade Estadual Paulista (Unesp)
Araraquara, São Paulo, Brazil
g.kundlatsch@unesp.br

Elibio Rech
Embrapa Genetic Resources and Biotechnology
Brasília, Distrito Federal, Brazil

Leonardo T. Duarte
School of Applied Sciences
Universidade Estadual de Campinas (Unicamp)
Limeira, São Paulo, Brazil

ABSTRACT

Intrinsic transcription terminators are central to the modularity and predictability of synthetic gene circuits. We leveraged a curated library of 582 bacterial terminators to train a predictive model and, from this surrogate, developed open source tools for terminator performance prediction and design. Each sequence was encoded by 130 sequence derived descriptors across four regions: A-tract, hairpin, loop, and U-tract. After performance based feature selection, 16 high impact attributes were retained to compare predictive models. A grid search optimized XGBoost model achieved the best average performance, exceeding the previously reported model as well as linear regression, MLPRegressor, and ensemble approaches. SHAP analysis demonstrated that U tract features indicate the importance of a more distal region than previously described and that the influence of initial hairpin GC content extends beyond the expected range. From the final model, we implemented two tools. The Terminator Strength Predictor computes features from an input sequence and returns a quantitative strength and a binary strong or weak class. Validation with experimentally characterized terminators from four bacteria, not represented in the training dataset, showed that the model reproduces relative efficiency rankings and assigns consistent classes. The Terminator Factory performs surrogate based optimization for target driven design under user defined strength and length constraints. It enabled enumeration of length specific sets of maximally strong terminators, design of synthetic terminators (TK and miniTK), and optimization of a native T7 terminator sequence. Designs were validated *in vivo* in *Escherichia coli* using the original library protocol. In addition, a new *in vitro* transcription assay based on fluorescent RNA aptamers (Broccoli and Mango III) was developed to further characterize the terminators. *In vivo*, TK exceeded the strongest reference terminator (Tmax) and miniTK showed high efficiency, while *in vitro* both showed the high performance. These results indicate that the model captures sequence to function relationships that support prediction with TerSP and design with TerFac.

1 INTRODUCTION

Transcription terminators are regulatory elements that mediate the dissociation of RNA polymerase from the DNA template and nascent RNA by destabilizing the elongation complex (EC), a stable intermediate in the transcription cycle (Ray-Soni et al., 2016). Terminators are classified as factor dependent or intrinsic, also known as Rho independent. Factor dependent terminators require trans acting proteins such as Rho and Mfd to induce EC collapse (Roberts & Park, 2004; Hao et al., 2021), whereas intrinsic terminators rely on their nucleotide sequence and secondary structure (You et al., 2023). Although intrinsic terminators function without dedicated factors, accessory proteins such as NusA and NusG can increase their strength (Mandell et al., 2022). Their simplicity, genetic compactness, and portability make them widely used in synthetic biology and metabolic engineering (Cui et al., 2021).

Efficient termination is essential for modular and predictable gene circuits. Incomplete termination causes transcriptional read through into downstream elements, leading to aberrant expression, circuit malfunction, and metabolic burden. Replacing a leaky T7 terminator with three intrinsic terminators increased volumetric protein yield in *E. coli* by 1.6 fold through improved termination and reduced resource competition (Mairhofer et al., 2015). In *Bacillus subtilis*, inclusion of a terminator increased gene expression by 20 fold relative to a terminator-less construct (Castillo-Hair et al., 2019). As construct complexity increases, stringent termination is required in densely packed transcriptional units, where high RNA polymerase flux across short intergenic regions promotes transcriptional interference. Multigene engineering also requires sequence-diverse terminators with predefined strengths to permit DNA synthesis and reduce homologous recombination arising from repeated elements (Chen et al., 2013). Here, we aim to expand the design space of synthetic terminators.

Despite their simplicity, intrinsic terminators remain difficult to rationally engineer or predict. Hoon et al. (De Hoon et al., 2005) proposed a decision rule for intrinsic terminator prediction in *B. subtilis* and other *Firmicutes*, enabling terminator classification and supporting conservation of intrinsic termination in this phylum. Kingsford et al. (Kingsford et al., 2007) identified low energy hairpins followed by a U tract and assigned termination likelihood scores, scanning a 4 megabase genome in under 50 s. Cambray et al. (Cambray et al., 2013) modeled cotranscriptional RNA folding using hairpin stability, GC content at the stem base, and U tail quality, but achieved high accuracy only after excluding 23 of 54 terminators and training on 31, indicating incomplete feature capture. Chen et al. 9 characterized 582 *E. coli* terminators and proposed a hybrid shearing model involving hairpin nucleation and U tract ratcheting, yet explained only 40% of variance after removing weak terminators and failed to produce universally stronger synthetic designs, highlighting context dependence. A recent machine learning classification model using sequence and thermodynamic features can distinguish strong and weak terminators to some extent (Zhai et al., 2022).

To expand predictive capacity and enable design, we trained a supervised machine learning model on the 582 terminators from Chen et al. (Chen et al., 2013), compared architectures, selected an optimal model, and applied it to synthetic terminator design validated against the original library *in vivo* and *in vitro*. We developed an *in vitro* fluorescence assay using fluorescent light up aptamers (FLAPs) to quantify termination, providing transcription specific readouts and eliminating sequence dependent translation effects (He et al., 2020). Finally, we implemented Terminator Strength Predictor (TerSP) and Terminator Factory (TerFac) to predict and generate strength specific terminators.

2 MATERIAL AND METHODS

2.1 FEATURE ENGINEERING AND SELECTION

The model was trained on the Chen terminator dataset (Chen et al., 2013). Feature engineering was performed in Python 3.11.11 (XGBoost v1.7.6, Scikit-learn v1.3.2, Pandas v2.2.2, NumPy v1.26.4, Matplotlib v3.10.6). Synthetic and natural libraries were merged and 130 descriptors were computed across A tract, hairpin, loop, and U tract regions. Features included nucleotide composition, structural descriptors, diversity and transition metrics, and positional scoring. Composition metrics quantified nucleotide and GC content across regions and positions. Structural descriptors included stem length, paired bases, initial GC count, thermodynamic stability, and loop length. Diversity metrics included Shannon entropy, state changes, and transition ratios. All features were MinMax

scaled to [0,1]. Feature impact was evaluated using train and test R2, retaining the 16 most informative descriptors.

2.2 HYPERPARAMETER OPTIMIZATION

A grid of candidate values was defined for five XGBoost parameters: n_estimators (2000, 4000, 6000, 8000, 10000), learning_rate (0.001, 0.0025, 0.005, 0.0075, 0.01), max_depth (2–6), min_child_weight (1, 3, 5), and subsample (0.50, 0.65, 0.80). A base XGBRegressor(objective='reg:squarederror') was fitted via GridSearchCV. Overfitting was assessed by discarding parameter combinations with mean train–test R^2 gaps above 0.35 or train R^2 above 0.9. The optimal configuration was n_estimators 4000, learning_rate 0.001, max_depth 5, min_child_weight 3, and subsample 0.65.

2.3 ALTERNATIVE MODEL EVALUATIONS

We evaluated neural networks, linear regression, and ensemble models. Neural networks were assessed using nine MLP architectures, from one layer with 10 or 50 neurons to four layers (50, 100, 100, 50), via five fold cross validation. For each architecture, MLPRegressor with fixed activation, solver, random state, tolerance, and max iterations was trained and evaluated using R2 and mean squared error, with fold means and standard deviations computed and ranked by mean test R2. Linear regression was evaluated by five fold cross validation with fold wise experimental versus predicted plots and mean R2 reporting. Ensemble modeling combined XGBoost and MLP in a VotingRegressor using seven weight pairs, evaluated by five fold cross validation, ranked by mean test R2, and visualized by train versus test R2. The final model was trained on the full dataset and serialized with joblib. Feature contributions were analyzed in the final model using SHAP (SHapley Additive exPlanations) within five fold cross validation. In each fold, a TreeExplainer computed SHAP values and interaction matrices for the held out test set, which were aggregated across folds. Global interpretation used three visualizations: a SHAP summary plot, a heatmap of mean absolute interaction strengths, and a waterfall plot for the test instance with the highest total SHAP contribution.

2.4 TERMINATOR STRENGTH PREDICTOR (TERSP)

The offline TerSP was implemented as a standalone Python script that loads the trained model with joblib.load, computes raw region specific descriptors, normalizes them to [0,1], and predicts Average Strength. Input terminator sequences are loaded, converted from DNA to RNA, uppercased, and validated for fixed A tract, U tract, stem, and loop lengths. Raw and normalized features and predictions are generated per sequence. An online version was implemented as a Flask web service that reuses the same feature calculation pipeline. User provided sequences are validated, concatenated into a full stem, converted into a feature DataFrame, and scored by the model. Predictions are reported numerically and classified as strong at values of 40 or above.

2.5 TERMINATOR FACTORY (TERFAC)

To restrict the Terminator Factory optimizer to biologically plausible parameter combinations, we first constructed discrete mapping tables for each sequence region. For loops, we enumerated all lengths $L = 3$–16 and GC counts from 0 to L, normalized the features, removed duplicates, and exported unique loop combinations. For fixed-length A tracts (8 nt) and U tracts (12 nt), we generated all possible sequences (65,536 and 16,777,216, respectively), computed and normalized regional features, deduplicated entries, and exported the resulting mappings.

Hairpin stems were treated compositionally, as brute-force enumeration becomes prohibitive at large L. For each even $L = 6$–48, with $\ell = L/2$, we enumerated all nucleotide count compositions whose A, C, G, and U counts sum to ℓ, enforcing that the second stem half is the reverse complement of the first. Feature bins were populated using an adaptive Monte Carlo scheme, sampling 5,000 random permutations per batch per composition, computing normalized tuples of state change, initial 5′ GC run length, and Shannon entropy, and adding novel bins. Sampling stopped after three consecutive batches without new bins, with convergence logs every 10,000 draws and parallel execution via ProcessPoolExecutor, totaling 2.33×10^8 iterations. Two deterministic prototypes per composition

(sorted and jagged) were added, mirrored by reverse complement, clamped to biological limits, and normalized to $[0, 1]$. Exhaustive enumeration for $L = 6$–24 produced identical normalized bins.

Optimization used SciPy differential_evolution with the best1bin strategy and Latin hypercube initialization on a continuous 16-feature vector, followed by snapping to the nearest discrete bins: loop features were rounded to loop_key and nearest GC; A- and U-tract features were snapped by nearest entropy, then minimum Euclidean distance over state change and composition; hairpin features were snapped by length, then initial GC, entropy, and state change. The objective returned the negative predicted strength and enforced a hairpin length cap. We used a population of 100, 3,000 generations, a tolerance of 1×10^{-5}, mutation rates of 0.7–1.1, recombination of 0.6, and five random seeds, retaining the best solution.

Sequence assembly used composition filters and selective enumeration, followed by concatenation. A Flask service exposed target-driven optimization with optional constraints via Redis RQ, redirecting to a status page with three-second polling and reporting predicted strength, denormalized features, generated sequences, and execution logs.

2.6 Experimental validation of terminator function

To evaluate terminator performance in a cellular context, candidate terminators were first characterized *in vivo* using a dual-fluorescent reporter system in *E. coli*. For assays, strains were grown in 4 milliliters LB with 100 micrograms per milliliter ampicillin at 37 degrees Celsius and 1000 rpm, induced with 10 milligrams per milliliter L arabinose for 4 hours, diluted in PBS, and analyzed on a CytoFLEX B5 R3 V5 cytometer. Sixty thousand events were recorded, live cells were gated by FSC A and SSC A, and fluorescence was quantified as geometric means using CytExpert.

To decouple terminator function from cellular regulation and transcriptional context, terminators were independently assessed *in vitro* using a fluorescence-based RNA aptamer transcription assay. Five fluorescent light up aptamers Corn, Broccoli, Mini Spinach, Mango II, and Mango III were evaluated as RNA reporters downstream of a T7 promoter. Complementary oligonucleotides were annealed by heating to 98 degrees Celsius and gradual cooling. After selecting the optimal pair, Broccoli and Mango III were combined in a single construct separated by candidate terminators, with structural adapters to ensure proper folding (Bains et al., 2023). DNA templates were synthesized as gene fragments, PCR amplified, purified, and quantified. In vitro transcription conditions were systematically optimized, including reaction volume, fluorophores, and ionic composition. Potassium and magnesium concentrations were adjusted to maximize fluorescence, and inorganic pyrophosphatase was added to prevent pyrophosphate accumulation. The final 10 microliter reaction combined Broccoli and Mango III aptamers with optimized buffer, salts, fluorophores, rNTPs, and T7 RNA polymerase.

3 Results

3.1 Feature engineering and model optimization

The Chen dataset of 582 intrinsic transcription terminators was used to evaluate sequence features describing termination mechanisms, using segmentation into A tract, hairpin, loop, and U tract. Over 130 candidate descriptors were assessed, and 16 were retained based on predictive contribution. The final six nucleotides of the A tract and the first six of the U tract were most informative, with adenine and uracil content outperforming metrics derived from broader or narrower regions, consistent with prior observations (Chen et al., 2013). Additional U tract patterns were observed, as uracil content across the first ten positions and cytosine content across the full region also improved performance, although dataset bias cannot be excluded. For the hairpin, the count of G or C nucleotides preceding the first A or U was more predictive than restricting analysis to the first three positions, suggesting an extended functional region to what was previously described. Free energy based features were excluded, considering structure as an emergent property from sequence, and sequence descriptors such as Shannon entropy and state change counts were prioritized, consistent with prior RNA modeling studies (Angenent-Mari et al., 2020).

Model optimization was performed using exhaustive grid search of XGBoost parameters (Friedman, 2001) while monitoring the gap between training and test R^2 values to avoid overfitting. Parame-

ter sets with gaps above 0.3 and training R^2 above 0.9 were excluded. The optimal configuration achieved a training R^2 of 0.897 and a test R^2 of 0.652 under five fold cross validation, substantially outperforming the kinetic model reported by Chen, which achieved an R^2 of 0.117. Linear regression performed poorly with a test R^2 of 0.363, confirming the need for non linear models. A multilayer perceptron achieved a training R^2 of 0.571 and test R^2 of 0.491, and ensemble approaches did not exceed XGBoost performance. SHAP analysis demonstrated that U tract and A tract features dominated model predictions, with low entropy and low state change poly U regions extending to position 10 strongly contributing to high strength. Feature interactions were sparse, supporting feature orthogonality and model interpretability.

3.2 De novo design and experimental validation

To evaluate the generative capacity of the model, we designed intrinsic terminators with predefined strengths. To isolate the contribution of newly identified hairpin features, the A tract and U tract were constrained to six adenines and eight uracils, respectively, while allowing free hairpin design. Under these conditions, the model generated TK, a 72 nucleotide terminator with a predicted strength of 275. To facilitate further use by reducing synthesis and cloning complexity, hairpin length was further restricted, yielding miniTK, a compact 44 nucleotide variant with a predicted strength of 255. For comparison, the strongest terminator in the original dataset had a predicted strength of 238.

Designed terminators were characterized in *E. coli* using the same dual reporter assay originally described by Chen. Benchmarking against B0010, a widely used terminator with reported strength 83, and L3S2P21 (the strongest terminator in the original library and here termed Tmax) confirmed assay consistency, as Tmax exhibited high termination efficiency relative to the no terminator control. In contrast to previous reports, B0010 performed similarly to Tmax. TK outperformed Tmax, while miniTK was classified as strong (Figure 1a). TK likely represents the strongest terminator tested in *E. coli*.

Since transcription and translation are coupled *in vivo* and influenced by DNA copy number, mRNA stability, and accessory factors such as NusA, NusG, and Rho (Sipos et al., 2007), we developed an *in vitro* transcription assay to isolate intrinsic termination. Multiple fluorescent light up aptamers were evaluated, and Broccoli was selected based on signal strength and compatibility with Mango III, with no detectable crosstalk. Single stranded templates performed poorly relative to double stranded templates and were therefore excluded. RNA insulators (Bains et al., 2023) were incorporated to prevent structural interference, and folding simulations confirmed compatibility. Reducing the NTP concentration from 5 mM to 0.3 mM minimized transcriptional readthrough and enabled clear discrimination between terminator strengths.

Using this optimized assay, Tmax outperformed T41 (the weakest of the strong terminators in the original library), validating the system. The synthetic terminator TK scored 1.72 and exceeded Tmax, whereas miniTK scored 1.12 and performed comparably. The wild type T7 terminator behaved as an intermediate strength terminator 7,20, while an optimized T7 variant incorporating model derived A tract and U tract sequences showed increased efficiency, demonstrating model generalization beyond the training data (Figure 1b).

3.3 Software Implementation and Cross Species Validation

The optimized model was implemented in two complementary tools and released as freely available web based and offline software. Terminator Strength Predictor (TerSP) is intended for sequence based strength prediction, whereas Terminator Factory (TerFac) uses the same model to support strength targeted *de novo* terminator design.

Given an input sequence, TerSP validates sequence length, computes the 16 defining features, predicts terminator strength, and classifies outputs as strong or weak using the ≥ 40 threshold defined by Chen. To assess generalization beyond the training data, TerSP was applied to intrinsic terminators from *Corynebacterium glutamicum* (Kang et al., 2022), *Vibrio natriegens* (Stukenberg et al., 2021), and *Halomonas bluephagenesis* (Xu et al., 2022) (Figure 2a, 2b and 2c). TerSP (Figure 2d) correctly classified all *C. glutamicum* terminators and largely reproduced their experimental strength ranking, identified the single strong terminator in *V. natriegens*, and showed moderate performance for *H. bluephagenysis*, supporting conserved determinants of intrinsic termination across species.

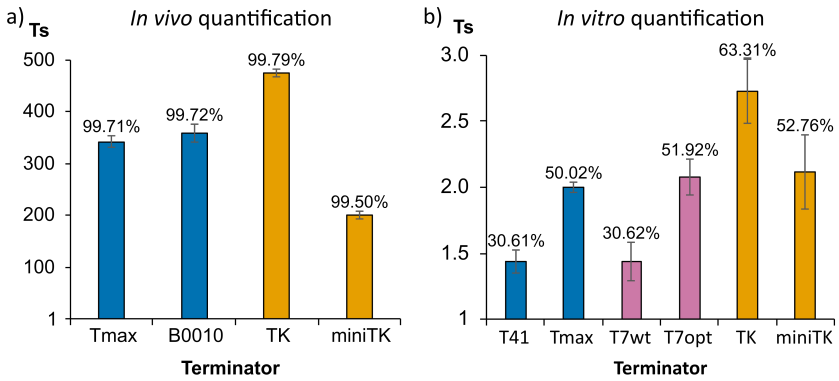

Figure 1: (a) Terminators were tested *in vivo* in the context of transcription, translation, and other cellular processes. The y axis shows terminator strength (Ts) values calculated according to the method described by Chen, while the values displayed above each bar indicate the corresponding converted termination efficiency (TE). (b) Terminators were tested *in vitro* in an isolated transcription system.

TerFac (Figure 2e), in turn, enables strength targeted design by restricting the optimization to discrete maps of biologically valid feature combinations. Loop, A tract, and U tract feature spaces were exhaustively enumerated, while hairpin stems were mapped by nucleotide composition under even length and reverse complement constraints, and populated by adaptive Monte Carlo sampling that matched exhaustive enumeration for short stems. During optimization, candidate solutions are snapped to valid feature bins, reconstructed into sequences consistent with the target features, and assembled into complete terminators satisfying user defined constraints. Analysis of optimized solutions revealed a preference for increased hairpin state changes at higher target strengths. Figure 3 provides an overview of model training and optimization, feasible features mapping, and the development of TerSP and TerFac.

4 Discussion and Conclusion

Machine learning can provide mechanistic insight in addition to predictive power. After optimizing predictive performance, we examined the most informative features. Since the relevance of conserved adenines upstream and uracils downstream of the hairpin have been reported for decades, it was unsurprising that low entropy, few state changes, and reduced incorporation of other nucleotides were associated with stronger termination. Within the A tract, the six nucleotides immediately upstream of the hairpin were most influential, consistent with previous work (Chen et al., 2013). By contrast, extending U tract features to include uracils at position +10 and cytosines at position +12 improved performance, indicating that distal positions can still modulate termination. Shorter loops were generally favored, consistent with more stable hairpin formation. In the stem, the initial contiguous G C feature was more informative than GC content across the first three positions, which had been highlighted previously. Hairpin length also emerged as a key determinant, consistent with established termination mechanisms.

Recently, Li et al. (Li et al., 2025) reported a terminator strength model trained on the same dataset. Their pipeline used k mer based descriptors and related encodings (Chen et al., 2014) that expand feature dimensionality by orders of magnitude and incorporated thermodynamic and secondary structure metrics that we intentionally omitted. Although their reported test R^2 exceeded ours, the synthetic terminators they generated did not outperform B0010, and the hairpin related insights identified here did not emerge, emphasizing how feature design and overfitting control shape both conclusions and generative utility on limited datasets.

We then applied our model to design synthetic terminators and validated them *in vivo* and *in vitro*. *In vivo* results matched predictions, with TK outperforming MiniTK and Tmax, demonstrating that the model can generate a novel high efficiency terminator. Although B0010 displayed higher in vivo

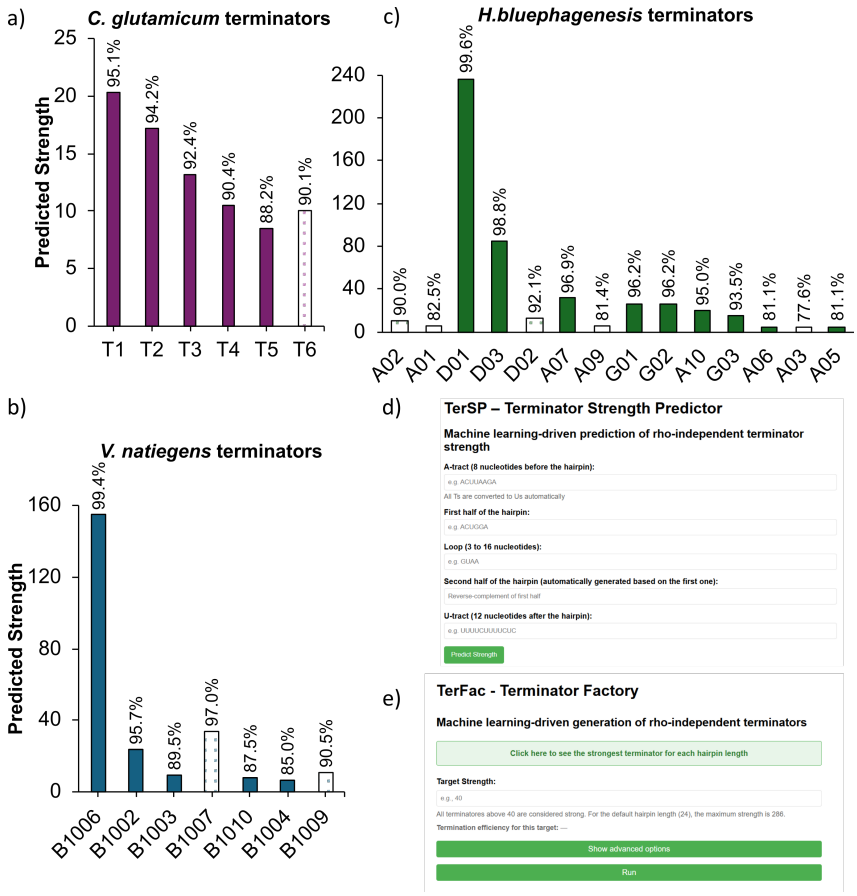

Figure 2: Terminator's strength prediction using TerSP beyond *E. coli*. The y-axis displays terminator-predicted strength values (Ts), while the values shown above each bar indicate the corresponding converted termination efficiency (TE) for (a) *C. glutamicum* , (b) *V. natriegens* and (c) *H. bluephagenesis*. Terminators are ordered by experimental strength, strongest to the left. Solid bars indicate correct ranking, while cross-hatched bars indicate incorrect predictions, overestimated in a and b, and underestimated in c. The graphical interface of TerSP (d) and TerFac is also presented (e).

efficiency than previously reported (Chen et al., 2013), the model still produced optimized designs, indicating robustness to potential dataset inconsistencies.

To isolate intrinsic termination from cellular context, we developed a dual aptamer in vitro assay using fluorescent light up aptamers. This assay confirmed TK as the strongest terminator and yielded lower absolute efficiencies than *in vivo*, consistent with enhancement by cellular metabolism. Relative to prior workflow, this approach eliminates cloning, shortens experimental time, facilitates automation, and improves reproducibility. Finally, we introduced two open source tools, TerSP and TerFac, that apply the trained model as a surrogate for experimentation. TerSP supports prediction beyond the training organism, whereas TerFac enables strength targeted design of sequence diverse terminators to tune circuits and reduce recombination risk. Together, these advances expand the synthetic biology toolkit and support construction of increasingly complex genetic systems.

5 Code Availability

All custom code used in this work, including that used to train and test machine learning models, and to generate analyses for different models and parameters, can be obtained at: https://github.com/gkundlatsch/TerSP-TerFac. Additionally, both software can be run using a browser from our Lab website https://taaqui.unesp.br/a3lf1 .

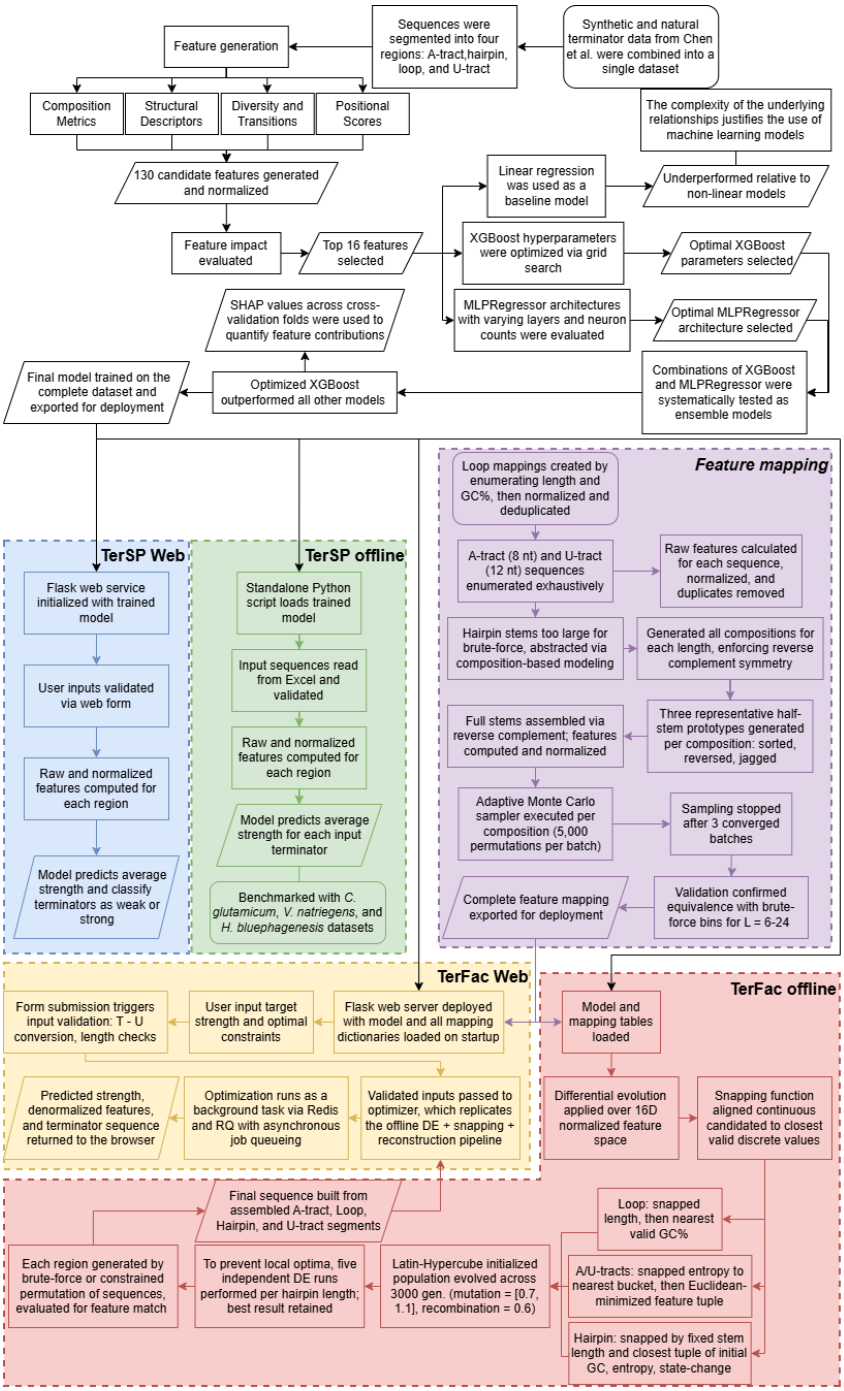

Figure 3: Development of TerSP and TerFac. The uncolored region outlines the model development process, including feature engineering, selection, and training. Colored areas correspond to the major components of the deployed system: TerSP Web (blue), TerSP offline (green), feature mapping (purple), TerFac Web (yellow), and TerFac offline (red).

REFERENCES

Nicolaas M Angenent-Mari, Alexander S Garruss, Luis R Soenksen, George Church, and James J Collins. A deep learning approach to programmable rna switches. *Nature communications*, 11(1):5057, 2020.

Jasleen Kaur Bains, Nusrat Shahin Qureshi, Betül Ceylan, Anna Wacker, and Harald Schwalbe. Cell-free transcription-translation system: a dual read-out assay to characterize riboswitch function. *Nucleic Acids Research*, 51(15):e82–e82, 2023.

Guillaume Cambray, Joao C Guimaraes, Vivek K Mutalik, Colin Lam, Quynh-Anh Mai, Tim Thimmaiah, James M Carothers, Adam P Arkin, and Drew Endy. Measurement and modeling of intrinsic transcription terminators. *Nucleic acids research*, 41(9):5139–5148, 2013.

Sebastian M Castillo-Hair, Elliot A Baerman, Masaya Fujita, Oleg A Igoshin, and Jeffrey J Tabor. Optogenetic control of bacillus subtilis gene expression. *Nature communications*, 10(1):3099, 2019.

Wei Chen, Tian-Yu Lei, Dian-Chuan Jin, Hao Lin, and Kuo-Chen Chou. Pseknc: a flexible web server for generating pseudo k-tuple nucleotide composition. *Analytical biochemistry*, 456:53–60, 2014.

Ying-Ja Chen, Peng Liu, Alec AK Nielsen, Jennifer AN Brophy, Kevin Clancy, Todd Peterson, and Christopher A Voigt. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature methods*, 10(7):659–664, 2013.

Wenjing Cui, Qiao Lin, Ruichun Hu, Laichuang Han, Zhongyi Cheng, Linpei Zhang, and Zhemin Zhou. Data-driven and in silico-assisted design of broad host-range minimal intrinsic terminators adapted for bacteria. *ACS Synthetic Biology*, 10(6):1438–1450, 2021.

Michiel JL De Hoon, Yuko Makita, Kenta Nakai, and Satoru Miyano. Prediction of transcriptional terminators in bacillus subtilis and related species. *PLoS computational biology*, 1(3):e25, 2005.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.

Zhitai Hao, Vladimir Svetlov, and Evgeny Nudler. Rho-dependent transcription termination: a revisionist view. *Transcription*, 12(4):171–181, 2021.

Zhiyun He, Yanting Duan, Weiji Zhai, Xiaomei Zhang, Jinsong Shi, Xiaojuan Zhang, and Zhenghong Xu. Evaluating terminator strength based on differentiating effects on transcription and translation. *ChemBioChem*, 21(14):2067–2072, 2020.

Dong Hun Kang, Sung Cheon Ko, Yu Been Heo, Hyun Jeong Lee, and Han Min Woo. Robomoclo: a robotics-assisted modular cloning framework for multiple gene assembly in biofoundry. *ACS Synthetic Biology*, 11 (3):1336–1348, 2022.

Carleton L Kingsford, Kunmi Ayanbule, and Steven L Salzberg. Rapid, accurate, computational discovery of rho-independent transcription terminators illuminates their relationship to dna uptake. *Genome biology*, 8 (2):R22, 2007.

Jie Li, Lin-Feng Wu, Kai Liu, and Bin-Guang Ma. Intelligent design of escherichia coli terminators by coupling prediction and generation models. *ACS Synthetic Biology*, 14(9):3744–3752, 2025.

Juergen Mairhofer, Alexander Wittwer, Monika Cserjan-Puschmann, and Gerald Striedner. Preventing t7 rna polymerase read-through transcription a synthetic termination signal capable of improving bioprocess stability. *ACS synthetic biology*, 4(3):265–273, 2015.

Zachary F Mandell, Dani Zemba, and Paul Babitzke. Factor-stimulated intrinsic termination: getting by with a little help from some friends. *Transcription*, 13(4-5):96–108, 2022.

Ananya Ray-Soni, Michael J Bellecourt, and Robert Landick. Mechanisms of bacterial transcription termination: all good things must end. *Annual review of biochemistry*, 85(1):319–347, 2016.

Jeffrey Roberts and Joo-Seop Park. Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Current opinion in microbiology*, 7(2):120–125, 2004.

Katalin Sipos, Réka Szigeti, Xiuzhu Dong, and Charles L Turnbough Jr. Systematic mutagenesis of the thymidine tract of the pyrbi attenuator and its effects on intrinsic transcription termination in escherichia coli. *Molecular microbiology*, 66(1):127–138, 2007.

Daniel Stukenberg, Tobias Hensel, Josef Hoff, Benjamin Daniel, René Inckemann, Jamie N Tedeschi, Franziska Nousch, and Georg Fritz. The marburg collection: A golden gate dna assembly framework for synthetic biology applications in vibrio natriegens. *ACS synthetic biology*, 10(8):1904–1919, 2021.

Mengmeng Xu, Yue Chang, Yuyan Zhang, Weizhe Wang, Jingyi Hong, Jiping Zhao, Xiaoyun Lu, and Dan Tan. Development and application of transcription terminators for polyhydroxylkanoates production in halophilic halomonas bluephagenesis td01. *Frontiers in Microbiology*, 13:941306, 2022.

Linlin You, Expery O Omollo, Chengzhi Yu, Rachel A Mooney, Jing Shi, Liqiang Shen, Xiaoxian Wu, Aijia Wen, Dingwei He, Yuan Zeng, et al. Structural basis for intrinsic transcription termination. *Nature*, 613 (7945):783–789, 2023.

Weiji Zhai, Yanting Duan, Xiaomei Zhang, Guoqiang Xu, Hui Li, Jinsong Shi, Zhenghong Xu, and Xiaojuan Zhang. Sequence and thermodynamic characteristics of terminators revealed by flowseq and the discrimination of terminators strength. *Synthetic and Systems Biotechnology*, 7(4):1046–1055, 2022.