Directed evolution of tunable bistable switch for cellular metabolic regulation

Team OUC-DE School of Marine Life, Ocean University of China, Qingdao, China

Abstract

Bistable switches are ubiquitous in natural and engineered biological systems, playing a key role in cellular regulation. We constructed a bistable switch consisting of tetracycline operon, arabinose operon, and lactose operon, allowing for the artificial control of the rate at which cells absorb metabolites from their environment and direct the flow of carbon within the cell. This aims to achieve rapid growth of engineered bacteria and efficient production of target metabolic products. We employed error-prone PCR and machine learning-assisted directed evolution to optimize the bistable system, increasing its tunability. Mutants were analyzed by observing fluorescence intensity and using visualization tools such as Pymol and molecular docking, to identify those that bind tightly to DNA and loosely to inhibitors. Eventually, we individually screened out an *araC* and *tetR* mutant that theoretically bind tighter with DNA and show better inhibition effect compared with the wild type.

Keywords： Bistable switches; Operon; Error-prone PCR; Machine learning-assisted directed evolution

Introduction

Bistability is a fundamental property of engineering and natural systems, with the ability to switch and maintain states. Prokaryotes couple the input and utilization of metabolites with feedback-regulated genetic circuits by constructing bistable switches, using inducible transcriptional and enzymatic components to create hybrid systems， which allow us to regulate the pathway easily[1]. All bistable switches developed so far, however, control the expression of target genes without access to other layers of the cellular machinery[2]. The binding of repressor proteins and operons can lead to incomplete pathway repression due to different selection and concentration of inducers, which affects the regulation of metabolic pathways by bistable switches[3]. Our experiment enhances the binding strength between the repressor protein and DNA by optimizing the switch[4], achieving lower basal level expression, induced high-level expression, and ultimately achieving bistable expression of the target gene.

This method has great potential to expand the functionality of biomolecular devices and can be used to separate growth and production processes during fermentation. Based on the specificity of hyaluronic acid produced by multiple metabolic pathways, we applied the designed bistable switch to regulate the carbon flow direction of engineering bacteria fermenting hyaluronic acid and tested the degree of performance improvement after optimizing the switch.

We attempted to construct a bistable switch using ordinary *Escherichia coli*. Due to the production of hyaluronic acid by the precursors UDP-glucuronic acid and UDP N-acetyl-glucosamine in the glucose metabolism pathway of *E. coli* under the action of *hasA*, we plan to use arabinose operons to regulate the *galU* and *udg* gene pathways in the UDP-glucuronic acid biosynthesis pathway of *E. coli*; In the biosynthesis pathway of UDP-N-acetyl-glucosamine, tetracycline operons are used to regulate gene pathways such as *glmU*, *glmS*, and *glmM*[5]. We plan to regulate the ratio of opening of the two pathways by changing the concentration of added inducers arabinose and dehydrated tetracycline, and during the growth process, make the metabolic pathway more inclined towards growth-related pathways, allowing the strain to grow rapidly first. During the fermentation

process, the metabolic pathway is biased towards the pathway that produces fermentation products, allowing the strain to produce a large amount of fermentation products. In the testing phase, we used GFP instead of the relevant genes in order to more accurately and quickly detect the performance of the mutants. After the design of the bistable switch was completed, we used error-prone PCR and computer simulation to evolve *araC* and *tetR* genes directly, to optimize the bistable switch, increase its ability to inhibit protein binding to DNA, reduce promoter leakage, and increase the performance of the bistable switch.

Materials and methods

Wet Lab

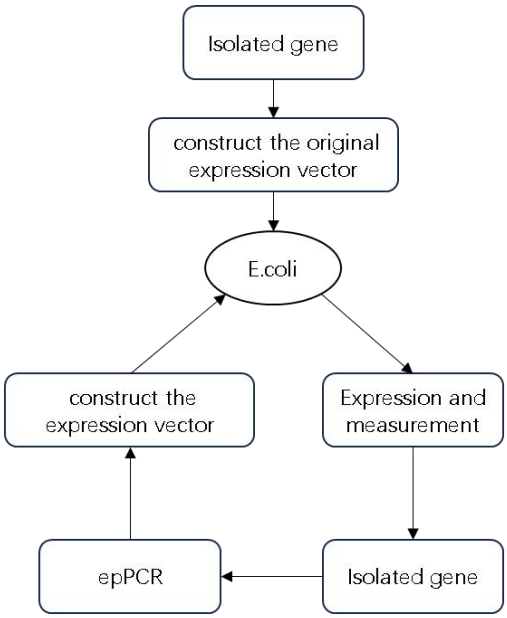

Figure 1 Flow of error-prone PCR screening for superior mutants using *E. coli*

Bacterial strains, media, chemicals and other materials

The bacterial strains and plasmids and others used in this study are listed in Supplementary Table 1.

E. coli T7-K12 was used for the construction of the original plasmid ptrc99a-araC and ptrc99a-tetR. The *araC* segment and *tetR* segment were mutated by epPCR to construct and screen a random mutation library.

E. coli bacteria were cultured in the Luria-Bertani (LB) medium for propagation. LB liquid medium was prepared by LB Broth, powder 4.8g and ampicillin 0.01 g in 0.1 liter of pure water, while LB solid medium contained 15 g/liter of agar.

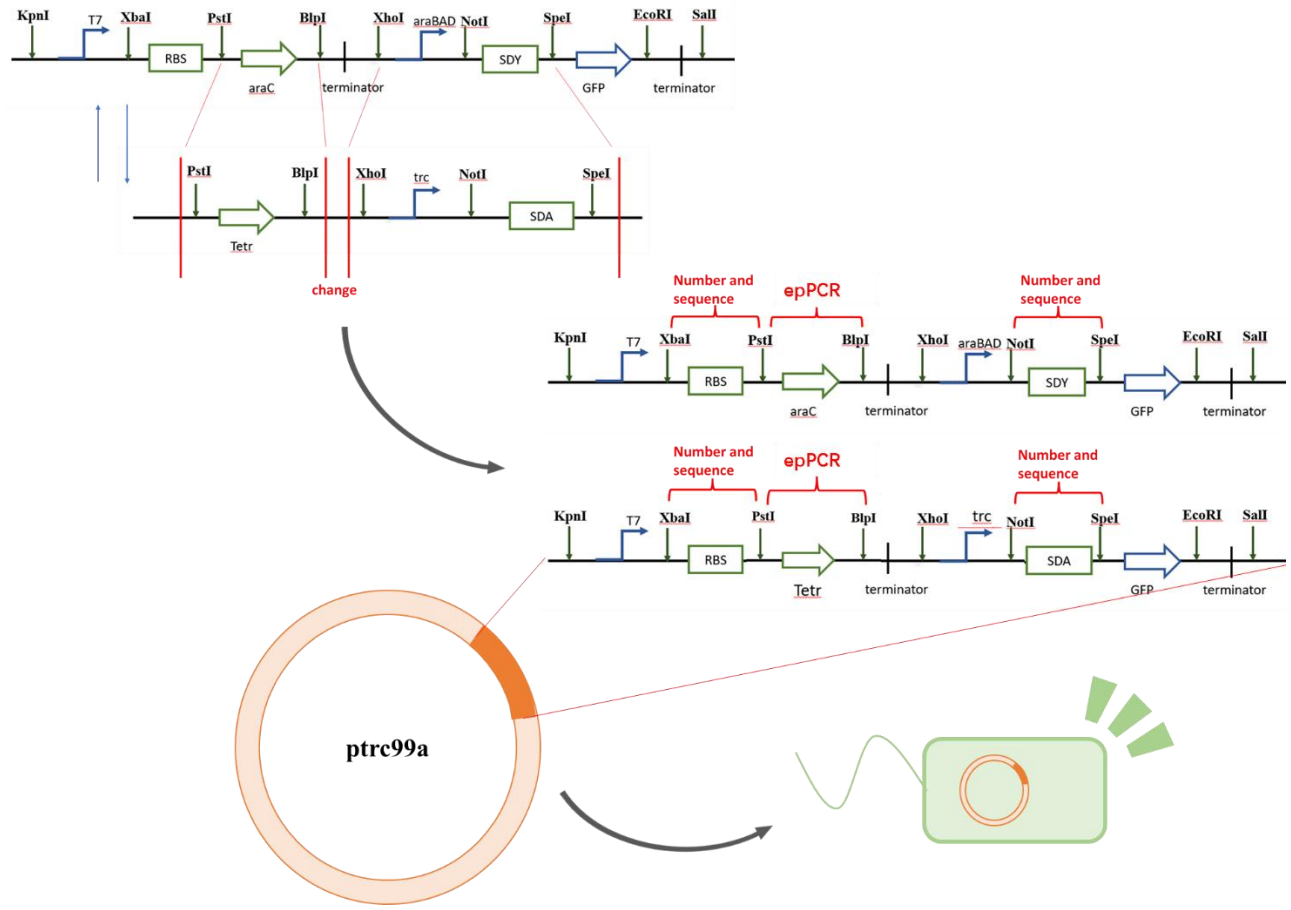

Figure 2 construct original plasmid ptrc99a-araC and ptrc99a-Tetr. We first synthesized a complete *araC*-containing *segment* and a partial *tetR*-containing gene *segment* for experiments, synthesized two *tetR*-containing *segment*s with complete functions by enzymatic cleavage, and finally introduced them into the ptrc99a plasmid to construct the original expression vector.

Expression vector construction

Analysis of the Multiple Cloning Site (MCS) of the ptrc99a segment vector sequence showed that the best double enzyme digestion sites on MCS were KpnI and SaII. We used snapGene to design the *araC* segment and the *tetR* segment. Ptrc99a was incised with KpnI and SaII, and the *araC* segment and *tetR* were connected to ptrc99a with T4 ligase, respectively. The sequence of *tetR*, *araC* synthesized by Sangon Biotech (Shanghai) are shown in supplementary Table 2. Primer sequence of *tetR*, *araC* synthesized by Sangon Biotech (Shanghai) are shown in supplementary Table 1.

The detailed steps are as follows:

Enzyme digestion of *araC* plasmid

First, we use double digestion to cut the two plasmids containing the target segment.

DNA	1μg
Kpn I	1μL
Sal I	1μL
10×LabFD^{TM} Buffer	2μL
ddH_2O	6μL

Table 1 Double digestion reaction system:

temperature	time
37°C	15min
80°C	20 min

Table 2 Reaction conditions

As the length of the *araC* target segment was similar to

that of the remaining portion of the original plasmid (both approximately 2500bp), we used a single digestion on *araC* to produce two segments of 1067bp and 1500bp. This allowed for clear separation when Rubber recovery.

DNA	1μg
Xho I	1μL
10×LabFD^{TM} Buffer	2μL
ddH_2O	8μL

Table 3 Single digestion reaction system

temperature	time
37°C	15min
80°C	20 min

Table 4 Reaction conditions

Enzyme digestion of *tetR* plasmid

DNA	1μg
Pst I	1μL
Spl I	1μL
10×LabFD^{TM} Buffer	2μL
ddH_2O	6μL

Table 5 Reaction system

temperature	time
37°C	15min
80°C	20 min

Table 6 Reaction conditions

Agarose gel electrophoresis

DNA	100 μL
6×Loading Buffer	20 μL

Table 7 System of Agarose gel electrophoresis

Rubber recovery

DNA segments were purifified from agarose gels by using the TIAgel Midi Purification Kit (TIANGEN, Bejing, China).

Digest ptrc99a

To faclitate segment interconnect, we use the same endonuclease to cut ptrc99a, *araC* and *tetR*.

DNA	1μg
Kpn I	1μL
Sal I	1μL
10×LabFD^{TM} Buffer	2μL
ddH_2O	6μL

Table 8 Double digestion reaction system:

temperature	time
37°C	15min
80°C	20 min

Table 9 Reaction conditions:

The carrier is connected to the *segment*

Escherichia coli DNA ligase and T4 DNA ligase are the main ligases used in genetic engineering. T4 ligase can catalyze the formation of phosphodiester bond between the 5 '-P end and 3' -OH end of double-stranded DNA, and also has good ligation efficiency for viscous end joining and end joining, it is wide application and have more users. Therefore, T4 ligase was selected for interconnection in this experiment.

The *araC segment* and *tetR* were connected to ptrc99a by T4 ligase.

Composition	Volume
Destination DNA segment	≈0.1pmol
Carrier DNA	≈0.01pmol
10×T4 DNA Ligase Buffer	1μL
T4 DNA Ligase	1μL
ddH_2O	Up to10μL

Table 10 Reaction condition

Overnight connection at 16℃, 3-5μL of the connection product was transformed into 100μL competent cells.

Observation

Fluorescence microscope was used to observe whether the cells were successfully imported.

Since there is green fluorescent protein in the arabinose promoter, it can appear green under 450-480nm blue light excitation. If green fluorescence is shown, it is considered to be successful introduction. Therefore, arabinose was added to the culture medium of cells containing the ptrc99a plasmid and

fluorescence was observed whether the color was developed.

Error-prone PCR

Error-prone PCR takes advantage of the fact that Taq DNA polymerase does not have a 3 '→5' proofreading function, and can introduce random mutations with a high probability under certain conditions.

The ptrc99a with target segment was mutated using the Controlled Error-prone PCR Kit, and the mutated plasmid was collected and purified by agarose gel electrophoresis and gel recovery.

Expected number of mutations	2	3	4	5	6	7	8
Amount of Mn（uL）	0	1.0	2.5	4.0	4.0	4.0	4.0
Amount of dG（uL）	0	0	0	1.0	2.0	3.0	4.0

Table 11 Error-prone PCR

ingredient	Amount of template 1	Amount of template 2	Amount of template 3
Error-prone PCR Mix, 10×(μL)	3	3	3
Error-prone PCR -specific dNTP,10×(μL)	3	3	3
Error-prone PCR-specific MnCl2	According to the above calculation		
Error-prone PCR -specific dGTP	According to the above calculation		
DNA template	1uL(1ng/uL)	1uL(10ng/uL)	1uL(100ng/uL)
PCR primer (10 uM each)	1 uL		
Error-prone PCR-specific	0.5μL		
Taq DNA polymerase			
Ultrapure water	Up to 30μL		

Table 12 Reaction system of Error-prone PCR

Initial denaturation	94°	3 minute
Error-prone PCR	94°	1 minute
	45°	1 minute
30-60 cycles		
Primer annealing	72°	1 minute

Table 13 Reaction condition of Error-prone PCR

Rubber recovery

Purified enzyme digested fragment, ligated fragments and plasmids. Repeated several times to increase concentration.

The error-prone PCR products were recovered from the gel, cloned and functionally screened or sequenced for single colonies, and the mutated DNA could be used as a template for the next round of error-prone PCR if needed to increase the mutation rate.

Dry Leb

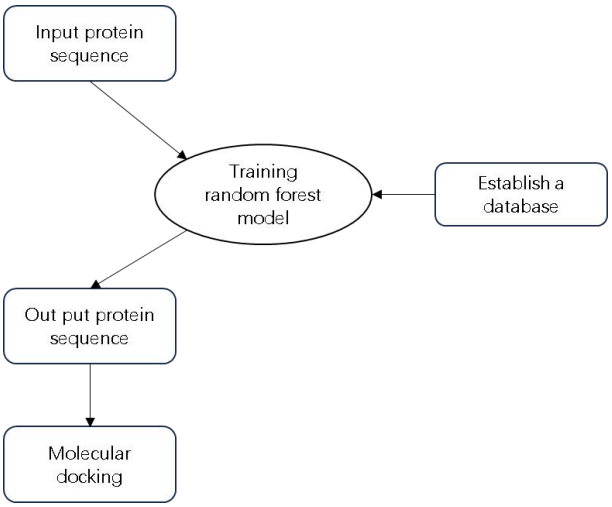

Figure 3 Computer-aided directed evolution process

Protein structure prediction-data processing

The database we used contains data on the binding of thousands of protein sequences and their amino acids to DNA. Firstly, we extracted the information of amino acids from the database which contained 20 common amino acids and encoded each amino acid, then stored

the information of each amino acid in a row vector. Each amino acid vector contained the types of the top n amino acids of the extracted amino acids, the types of the last n amino acids, the types of amino acids at the position, that position of the amino acids in the protein, the total number of amino acids in the protein, and information on the binding of amino acids at that position to DNA. Finally, we constructed a dataset containing 296610 amino acid vectors, each representing an amino acid and its information in the protein. (n=10)

Prediction and scoring calculation

After constructing the dataset, we used the first 200000 pieces as the training set and the last 100000 pieces as the prediction set. We used the random forest algorithm to calculate the relationship between DNA binding sites and other information, established a prediction model, and evaluated the overall probability of DNA binding. Using the protein sequence to be tested for extraction and prediction, a matrix containing all amino acid vectors of the protein will be generated. The prediction results will be filled in the last row of each column, and the DNA binding site data in the last row will be directly output to determine whether the protein has a DNA binding site. The evaluated binding probability will be summed up and output to represent the overall binding of the protein to DNA. The dataset is listed in Supplementary Figure 1.

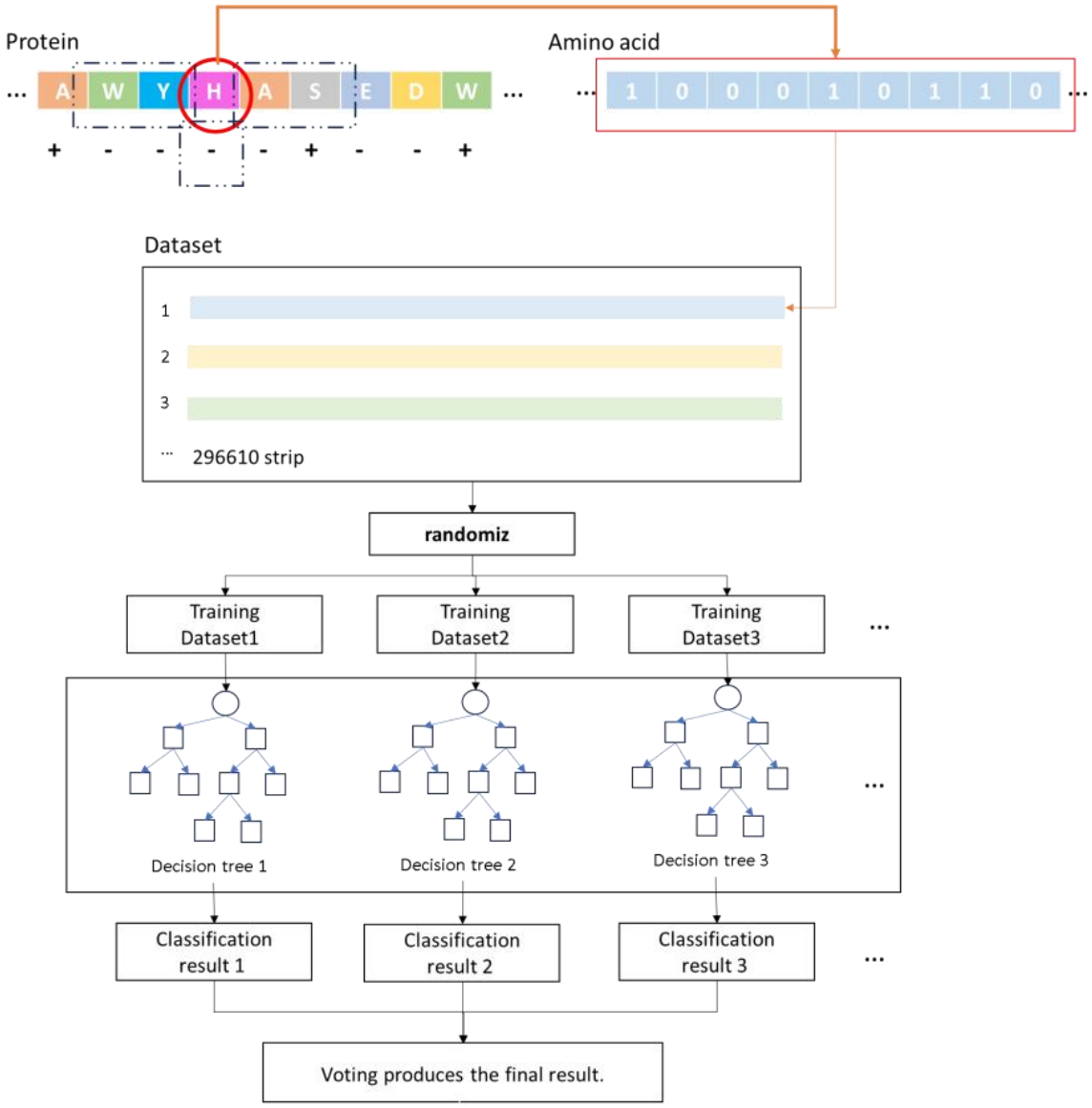

Figure 4 How to calculate the score P

Iteration process

We constructed an iterative model for protein directed mutation based on the prediction of the above data in the iterationprocess (figure 5). Firstly, input the amino acid sequence of the protein to be measured, evaluated its binding site and binding probability Pi by using a trained model, and then stored it. Randomly replaced the amino acids within the specified range of the protein and evaluated the binding probability Pi+1 after mutation. If the scores increased, the replaced amino acid sequence and its scores would be stored and be replaced in the next round. If the scores decreased or remained , the current scores and sequence would not be stored, and the amino acid sequence before replacement would be used for the next round of replacement. When the scores of 20 mutations did not increase, it was judged that it had reached a local optimal solution. Due to the fact that this algorithm could not specify the binding site of DNA, in order to avoid making significant changes to the protein active site, we limited its iteration frequency to no more than 1000 times to prevent excessive deviation of the binding site.

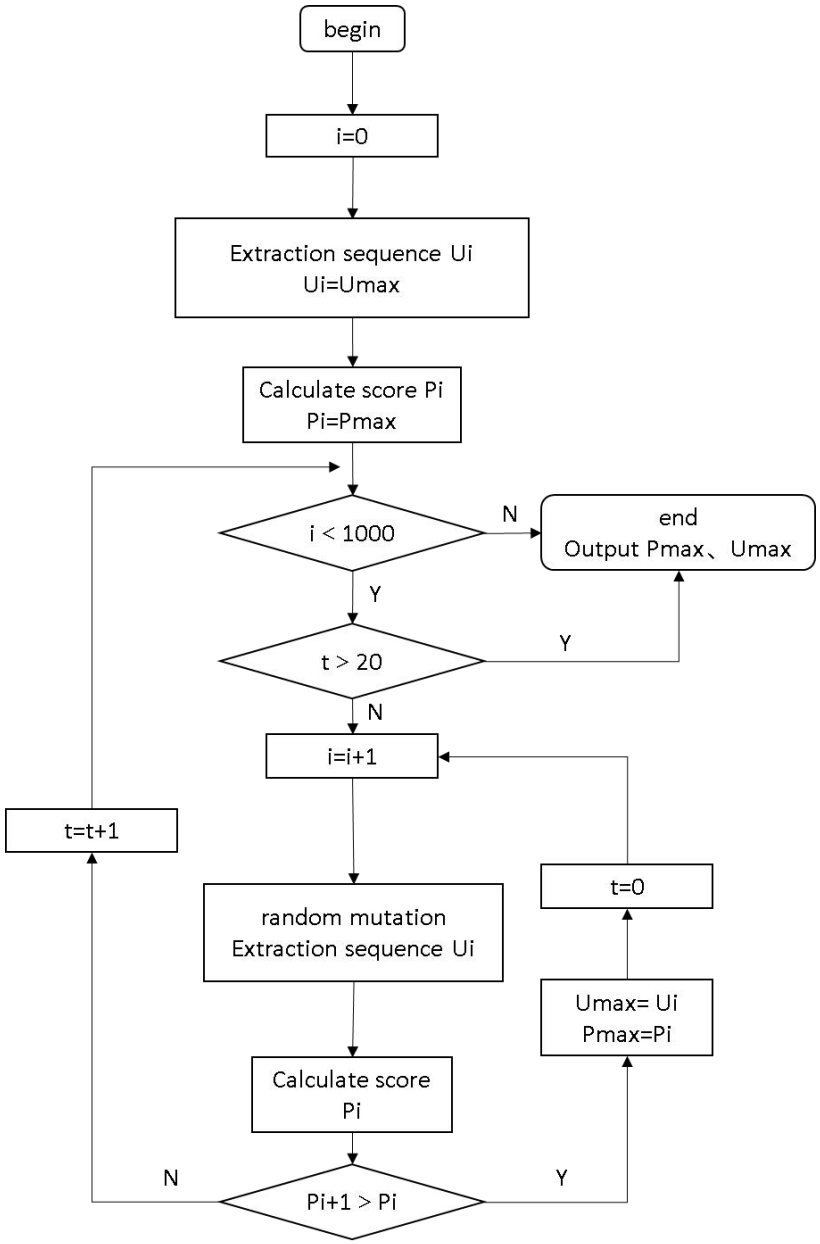

Figure 5 Program iteration process

Result

Structural simulation

AraC is composed of 292 amino acid residues and uses the active sites formed by the amino acids 198-219 and 246-269 to bind to DNA. The isolated AraC protein binds to araO1(-100 to -144), acting as an inhibitor. When the C protein binds with the inducer AraBAD, forming a complex, it binds to the araI region (-40 to -78) allowing RNA Pol to bind to the PBAD site (+140) and transcribe downstream genes.

TetR monomer is made up of 198 amino acid residues and, when dimerized, is folded by 10 alpha-helices connected by turns and loops. The three-dimensional structure of the TetR monomer is mainly stabilized by hydrophobic inverse helices. The sequence of amino acids 33-52 located at the N-terminal is the DNA binding region, composed of helices a1, a2, a3, and their symmetric helices a1', a2', a3'. The regulatory core, composed of helices a5, a10 and their symmetric helices a5', a10', controls dimerization and the binding sites of each monomer in the presence of divalent cations. Helices a5, a8, a10, and their corresponding helices form the core scaffold, and their structure is the most conserved in the entire TetR conformation.

To avoid significant protein structural changes, we

only subjected the DNA binding sites of AraC and TetR to computer-simulated random mutations. The optimized proteins obtained showed a slight change in their DNA binding site structures through structural simulations, while the rest of the structures remained largely unchanged. (Figure 6, Figure 7).

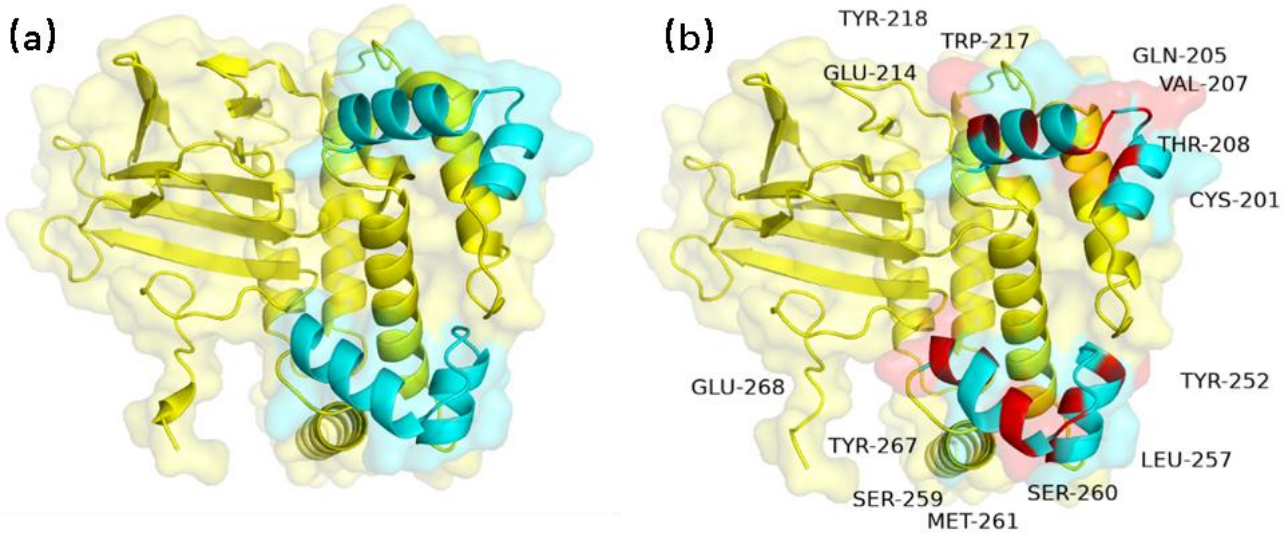

Figure 8 Comparison of AraC homologous modeling results. **(a)** Pre-mutation AraC structure **(b)** Homologous modeling structure of post-mutation AraC. AraC rated P from 12.98 to 15.94 . The blue part is the DNA-binding site of the protein, the yellow part is the rest of the structure, and the red part represents the amino acid after the mutation.

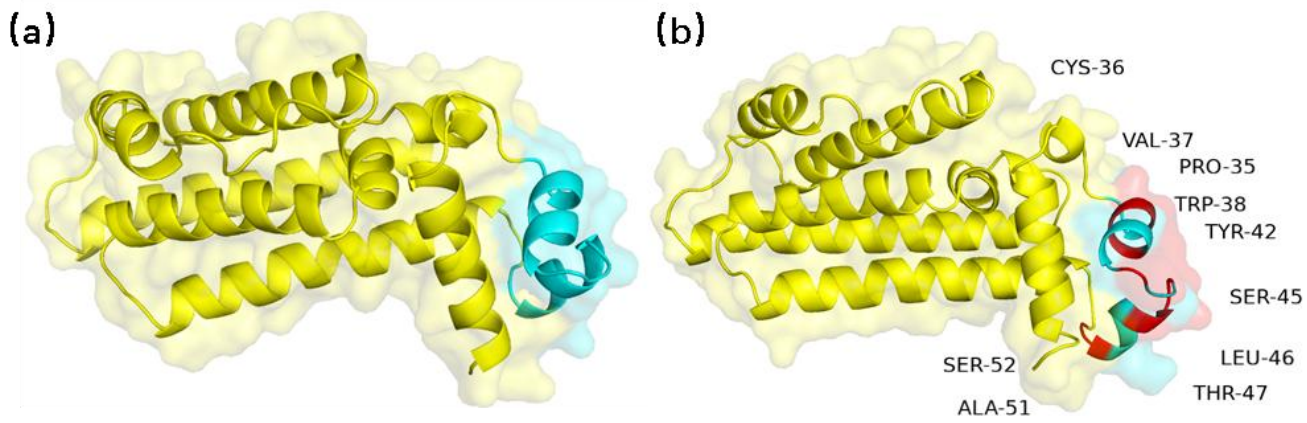

Figure 9 Comparison of TerR homologous modeling results. **(a)** Pre-mutation TetR structure (b) Post-mutation TetR homologous modeling structure. The score P of TetR ranges from 5.38 to 8.19. The blue part is the DNA-binding site of the protein, the yellow part is the rest of the structure, and the red part represents the amino acid after the mutation.

molecular docking

To validate the performance of the protein post-mutation, molecular docking was conducted using the simulated structures of both pre- and post-mutation AraC proteins with the araBAD promoter at positions (-144, -100) (Figure 5). Similarly, docking was performed using the simulated structures of pre- and post-mutation TetR proteins with specific sites on the tetracycline regulatory promoter (Figure 6). The results showed that The mutant has lower docking parameters

for DNA compared to the wild type and higher binding free energy for its deterrents. This suggests a tighter binding to DNA and a lower tendency to bind to its deterrents. This suggests a theoretical reduction in manipulator leakage and a smoothing of the change curve in the concentration of deterrents required for switching on, increasing its tunability. A comparison revealed a shift in the binding position of AraC post-mutation, although the movement was minor and is anticipated not to impact the actual effect. This confirms the effectiveness of the model, offering a rapid and efficient method to predict and improve DNA-binding proteins. Detailed docking data has been saved in the supplementary materials.

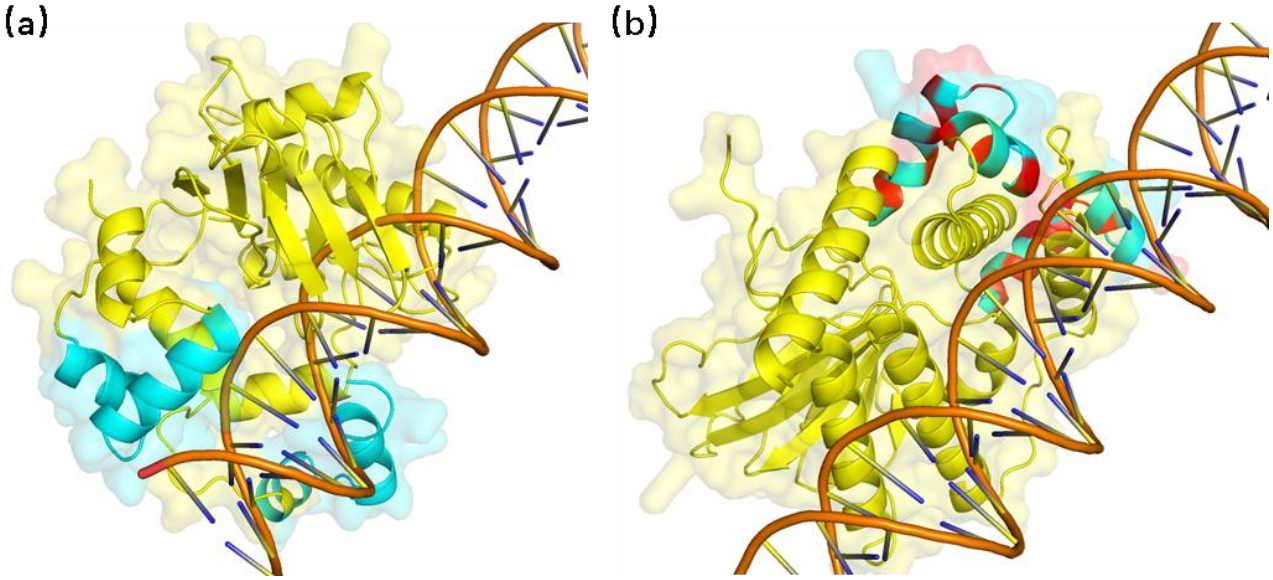

Figure 10 AraC molecular docking with DNA. **(a)** AraC and DNA docking results before mutation **(b)** AraC and DNA docking results after mutation. The blue part is the DNA-binding site of the protein, the yellow part is the rest of the structure, and the red part represents the amino acid after the mutation.

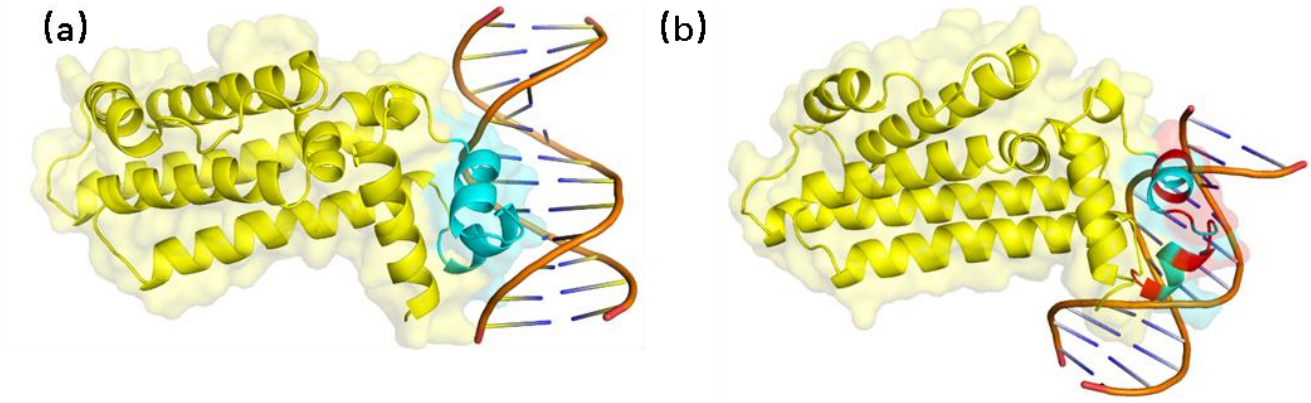

Figure 11 TetR molecular docking with DNA. **(a)** TetR and DNA docking results before mutation **(b)** TetR and DNA docking results after mutation. The blue part is the DNA-binding site of the protein, the yellow part is the rest of the structure, and the red part represents the amino acid after the mutation.

	Docking Score	Confidence Score	Ligand rmsd (Å)

AraC---DNA	-212.02	0.7756	93.36
TetR---DNA	-214.65	0.7847	45.55
AraCpro---DNA	-222.48	0.8099	97.96
TetRpro---DNA	-273.88	0.9226	95.74

Table 14 Docking data ： Docking Score: The docking energy scores.. Confidence Score： Credibility of docking. Ligand rmsd (Å): The ligand RMSDs from the input structures or modeled structures by homology modeling.

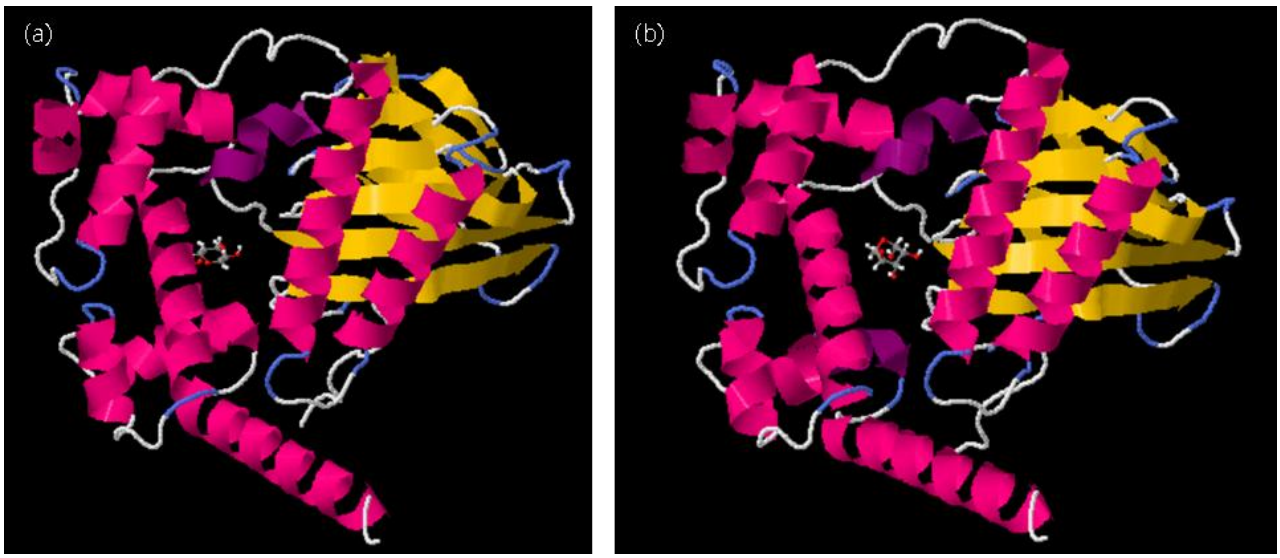

Figure 12 AraC molecular docking with arabinose. **(a)** AraC and arabinose docking results before mutation **(b)** AraC and arabinose docking results after mutation.

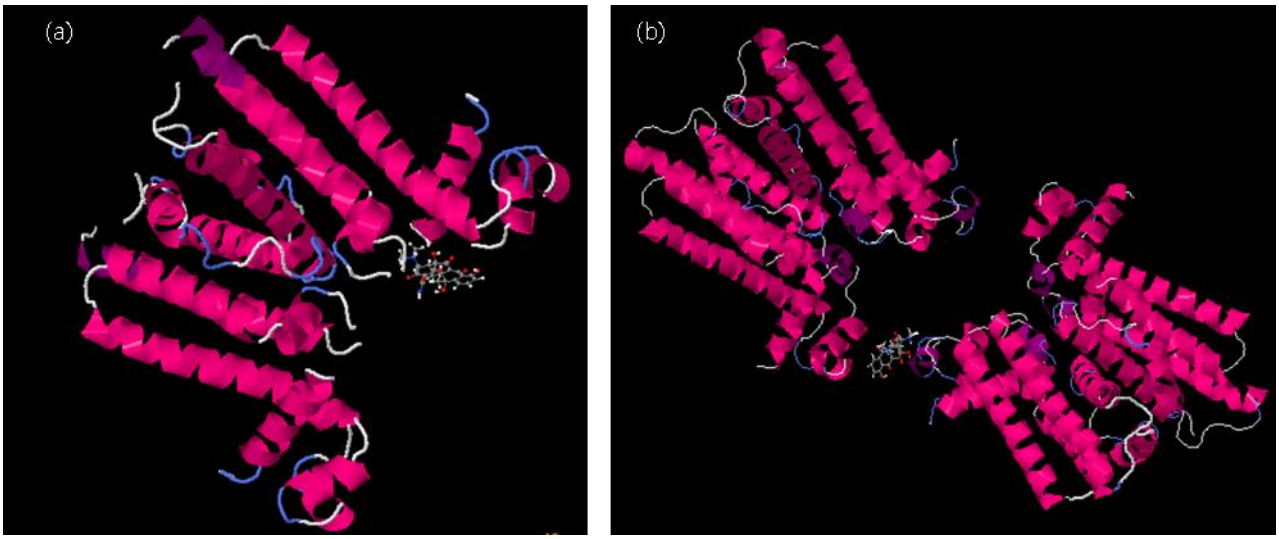

Figure 13 TetR molecular docking with tetracycline. **(a)** TetR and tetracycline docking results before mutation **(b)** TetR and tetracycline docking results after mutation.

	Estimated ΔG (kcal/mol)	FullFitness (kcal/mol)
AraC---arabinose	-6.79	-1318.70
TetR---tetracycline	-7.70	-2811.57
AraCpro---arabinose	-6.27	-1305.61

TetRpro---tetracycline	-7.45	-3908.49

Table 15 Docking data： Estimated ΔG Predicted binding free energy. FullFitness Usually used to describe the overall interaction energy between two molecules, TetRpro because it is docked using the tetramer model, FullFitness may be large.

Discussion

We have built a bistable switch to control the rate at which cells take up a metabolite from the environment [2] and are preparing to mutate the *araC* and *tetR* genes responsible for regulating the switch. Our goal is to screen for the most effective and sensitive switches by detecting downstream toxic proteins and fluorescent signals to aid in different applications of bistable switches(figure 7).

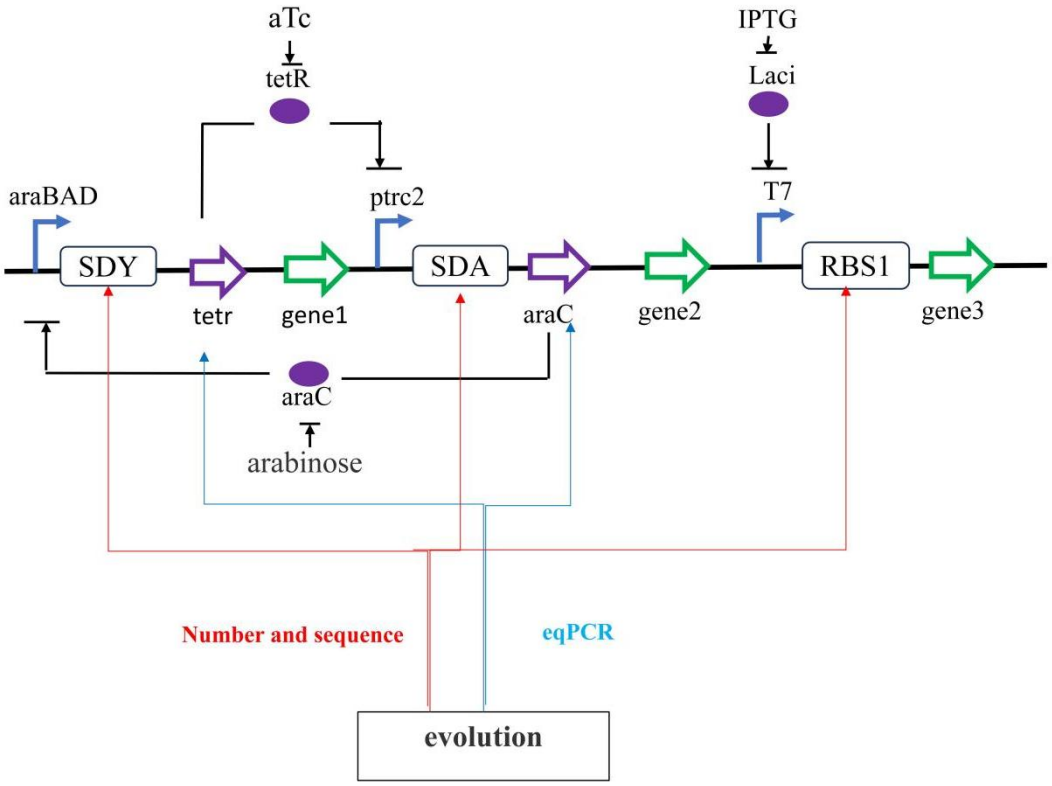

Figure 14 Gene regulatory switch. Act as a sign of evolution gene3 can be replaced with GFP . The change of gene1, gene2 and gene3 can be used to regulate various metabolic production pathways.Directed evolution of the ribosome binding site may also be one of the ways to optimize the switch

Other Researchers have knocked out genes such as *pfkA*, *pfkB*, *zwf*, *galR*, and *galS* in *E. coli*, and introduced *galU* and *kfiD* genes to enable them to activate the galactose pathway and reduce the amount of glucose entering respiratory metabolism, enabling them to efficiently utilize glucose and galactose to synthesize the rate limiting product UDP-glucuronic acid in the hyaluronic acid production pathway, thereby efficiently synthesizing hyaluronic acid. [5] Our designed switch aims to further optimize this strategy. We will conduct experiments in *E. coli* producing HA to achieve artificial control between the cell's growth and production modes.

We have spent a lot of time modifying *E. coli* k12 to

enable it to undergo the next step of directed evolution. However, we did not successfully transfer the designed plasmid into *E. coli* K 12. Many times of DNA agarose gel recovery, digestion and reconnection have made the concentration of our target DNA *segment*s too low. However, after computer prediction, we believe that the introduction of gene regulation switches into the directed evolution process may help the biosynthesis of *E. coli*.

The algorithm we constructed for predicting whether a protein is a DNA binding protein and evaluating its binding strength, and used it to simulate mutations in AraC and TetR proteins to improve protein DNA binding efficiency, reduce switch leakage, and increase switch performance.

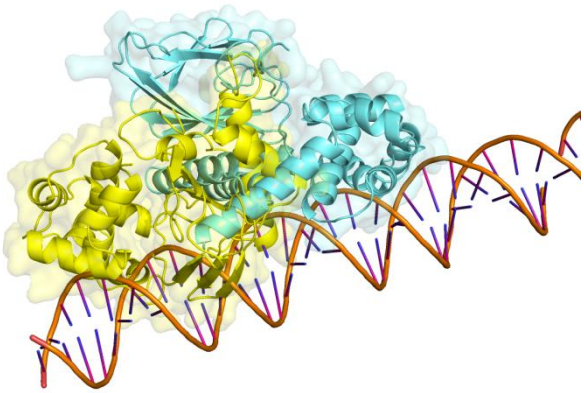

Figure 15 AraC-DNA binding site before (yellow) and after (blue) mutation.

Since the algorithm cannot specify the site where the protein binds DNA, certain changes to the DNA binding site occur before and after AraC mutation (Figure 8), so the algorithm needs to control the number of mutation sites to reduce the influence of binding site movement. Although the AraC and DNA binding sites are changed, they are still in the region of the operon-binding repressor protein, so the model can be considered valid

Future Directions

Our current major advancements come from the results of dry lab experiments. In the next step, we plan to express the proteins predicted in E.coli to test their functionality, which will further validate the feasibility of the prediction model we have established..

Subsequently, we will use mathematical models and analysis to determine the parameters of the structure generating bistable, and verify the adjustable parameters in combination with wet experiments to obtain the optimal bistable switch design.

Conclusion

Although bistable systems are commonly found in cells, it is not until the rise of synthetic biology in recent years that the study of bistable systems has become popular. Bistable systems can be used not only for metabolic regulation, but also as components of genetic circuits such as biosensors and biocounters. The model of directed evolution provides new ideas for optimizing bistable systems, although further research is needed in this area in order to have a significant impact on the function of bistable systems.

Acknowledgments

Throughout the writing of this dissertation we have received a great deal of support and assistance.

We would first like to thank our supervisor, Prof. Guanpin Yang whose expertise was invaluable in formulating the research questions and methodology. Your insights made our research problems more complete, and you have given us a lot of care and help in every important link during the research.

In particular, We would like to thank Dr. Li Guo and graduate students in the lab,who provide valuable assistance with the laboratory and field tests.

We would also like to thank our team members, for their wonderful collaboration and rigorous humility working attitude.

We would particularly like to acknowledge our friendly classmates, the 2022 iDEC team, OUC_China, for sharing their experiences and always willing to go out of their way to answer our questions.

Finally,We also would like to express our sincere gratitude to our college--College of Marine Life Science, Ocean University of China, who always give us material and spiritual support for our experimental exploration.

Supplementary

araC sequence :
plasmid number: 230726HV8303-8
araC forward primer:
TCTGCAGAATGGCTGAAGCGCAAAATGATC
tetR reversed primer:
GTTATGCTAGTTATTGCTCAGCGGTGGCA
company: Sangon Biotech

tetR sequence:
plasmid number:230723PT6768-2
tetR forward primer:
GCGGCCTGCAGAATGTCCAGATTAG
tetr reversed primer:
GCTCAGCGGTGGCAGCAGCCAACTCA
company:Sangon Biotech

Shuffle T7-K12 competent cell:
Genotype: F′ lac, pro, lacIQ /ρ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1ρ(phoA) PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC (Spec^R, lacIq)ρtrxB rpsL150(Str^R)ρgorρ(malF)3
Company:Biomed

Enzyme:
The company of KpnI, XbaI , PstI, XhoI, NotI, SpeI, EcoRI, SalI: LABLEAD
The company of BlpI: Thermo SCIENTIFIC

The dataset:
http://webs.iiitd.edu.in/raghava/[6]

Reference

[1] Huang D, Holtz WJ, Maharbiz MM. A genetic bistable switch utilizing nonlinear protein degradation. J Biol Eng. 2012 Jul 9;6(1):9. https://doi.org/10.1186/1754-1611-6-9 PMID: 22776405; PMCID: PMC3439342.

[2] Oyarzún DA, Chaves M. Design of a bistable switch to control cellular uptake. J R Soc Interface. 2015 Dec 6;12(113):20150618. https://doi.org/10.1098/rsif.2015.0618 PMID: 26674196; PMCID: PMC4707844.

[3] Rolf Lutz, Hermann Bujard, Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements, Nucleic Acids Research, Volume 25, Issue 6, 1 March 1997, Pages 1203–1210, https://doi.org/10.1093/nar/25.6.1203

[4] von Hippel PH, Revzin A, Gross CA, Wang AC. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proc Natl Acad Sci U S A. 1974 Dec;71(12):4808-12. https://doi.org/10.1073/pnas.71.12.4808 PMID: 4612528; PMCID: PMC433986.

[5] Woo JE, Seong HJ, Lee SY, Jang YS. Metabolic Engineering of Escherichia coli for the Production of Hyaluronic Acid From Glucose and Galactose. Front Bioeng Biotechnol. 2019 Nov 21;7:351. https://doi.org/10.3389/fbioe.2019.00351 PMID: 31824939; PMCID: PMC6881274.

[6] Patiyal S, Dhall A, Raghava GPS. A deep learning-based method for the prediction of DNA interacting residues in a protein. Brief Bioinform. 2022 Sep 20;23(5):bbac322. https://doi.org/10.1093/bib/bbac322 PMID: 35943134.