Week1 (6.1-6.7): pYB1s-ndmDtB Plasmid Construction

(A) pYB1s-ndmDtB plasmid construction:

1. Enzymatic digestion

Enzymatic cleavage of pYB1s vectors using specific restriction endonucleases to obtain the desired pYB1s vector fragments.

2.(1) Amplification of P107 fragment by PCR

The target fragments were obtained by PCR. The RCR system was as follows:

PCR system (50μL)	
2×Mix	25ul
P107-F	2ul
P107-R	2ul
template	50ng
DDW	20ul

PCR		
98° C	5min	
98° C	30s	×25
59° C	30s	×25
72° C	15s	×25
72° C	5min	
25° C	∞	

2.(2) Amplification of NdmB fragments by PCR

The target fragments were obtained by PCR. The RCR system was as follows:

PCR system (50μL)	
2×Mix	25ul
ndmB-F	2ul
ndmB-R	2ul
template	50ng
DDW	20ul

PCR		
98° C	5min	
98° C	30s	×25
56° C	30s	×25
72° C	33s	×25
72° C	5min	
25° C	∞	

2.(3) Amplification of NdmDt fragments by PCR

The target fragments were obtained by PCR, and the RCR system was as follows:

PCR system (50μL)	
2×Mix	25ul
ndmDt-F	2ul
ndmDt-R	2ul
template	50ng
DDW	20ul

PCR		
98° C	5min	
98° C	30s	×25
54° C	30s	×25
72° C	10s	×25
72° C	5min	
25° C	∞	

3.Cutting glue recycling

These fragments were recovered from the gel using a cut-gel recovery method to reduce interference from the original template.

4. Connecting the four fragments using the Gibsion assembly method

Use the Gibson C116 method to join the above four fragments together

Gibson system	
0.02 x 2900bp	ng
0.04 x 1118bp	ng
0.04 x 407bp	ng
0.04 x 340bp	ng
2×c116 Mix	5ul
DDW	to 10ul

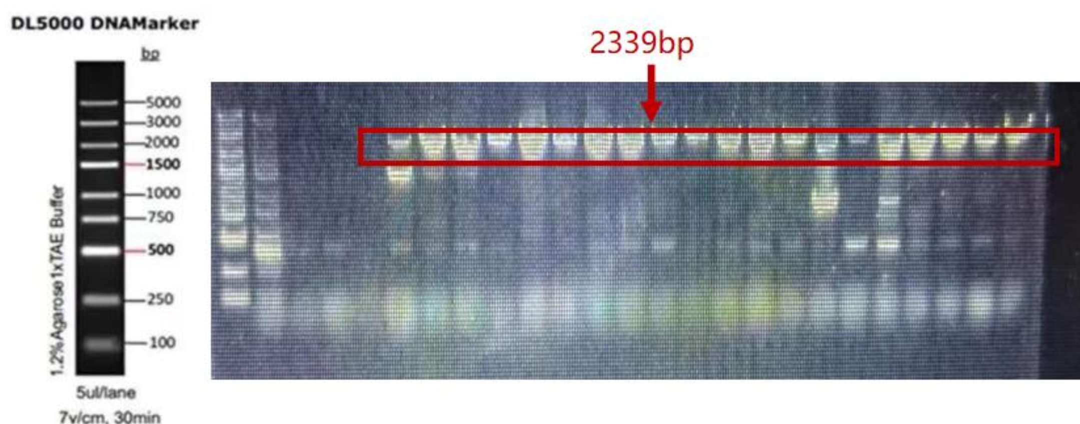
Gibson.	
50° C	30min
4° C	∞

5. Chemical transformation

Using the chemotaxis method, the ligation product was transformed into DH5α receptor cells and then spread on LB plates containing Amp resistance and incubated overnight at 37°C in inverted mode.

6. Colony PCR

PBF/PBR was selected as the primer, and the length of the target band should be 2339 bp, which was detected by agarose gel electrophoresis to confirm whether the expected fragment was obtained.

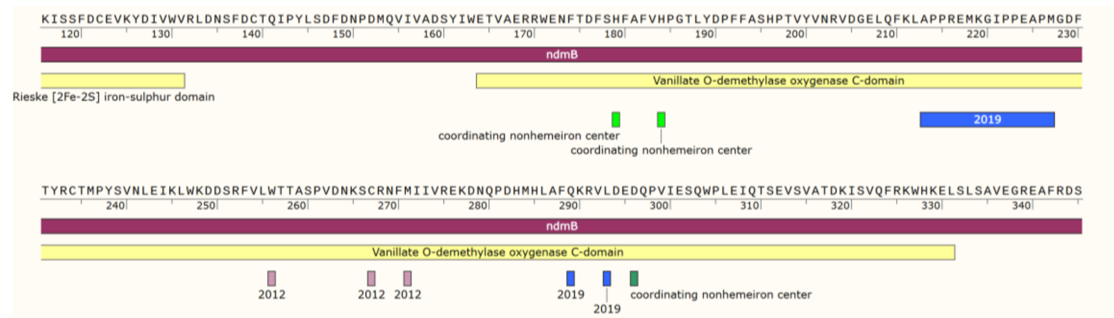


7. Extract plasmid, send for sequencing

Extract the plasmid, sequence it, and wait for the sequencing results.

Week2(6.8-6.14): Construction and screening of pYB1s-ndmDtBs mutant library

(i) Construction of a pYB1s-ndmDtBs mutant library:



1. Selection of mutation sites

Based on previous studies on NdmB, there are two sets of potentially critical combinatorial sites

① Trp256+Cys267+Met271

② Gln289+Leu293

Combinatorial random mutagenesis of the two sets of loci was performed separately, and DePX was used to screen for viable strains.

2. PCR amplification to obtain the target fragment

Using pYB1s-ndmDtB as a template (template plus 20~50ng/50μL), design and synthesise parsimonious primers, use high-fidelity enzymes, perform PCR (25 cycles to reduce unwanted mutations), and amplify 3~4 tubes.

3. Product purification

Subsequently, PCR product purification was performed, and then 200-300 ng were taken for DpnI digestion.

4. Connection, transformation

The purified fragments were Golden gate ligated and subsequently transferred to E. coli DH5 α sensing state and cultured at 37°C for 12~16h, as large as possible.

5. Wash with water, extract plasmid

Subsequently, the colonies were rinsed with water to extract the plasmid and the extracted plasmid was transformed into E. coli BW25113.

6. Screening

The transformed strains were spread on selective plates and the monoclonal strains were picked and transferred to deep-well plates in 800 μL of ZY medium + arabinose + IPTG + S50. The strains in the first four wells were used as non-mutated control strains, and all the strains were induced for 18 hours at 25°C. After incubation, the plates were centrifuged (4200 rpm, 10 min) and the supernatant discarded. After incubation, the plates were centrifuged (4200 rpm, 10 min), the supernatant discarded, and 400 μL of 1X M9 medium + 0.5 mM caffeine (equivalent to doubling the concentration of the organisms) was added. Whole-cell catalysis was continued at 25°C for 18 hours. After centrifugation, transfer 200 μL of supernatant to a 96-well plate and add 10 μL of supplemental solution (containing glucose, calcium chloride, magnesium chloride, trace inducer and antibiotics). Then add 10 μL of induced pYB1s-ndmDCEA (taking into account that it may grow slowly, the 96-well plate is sealed in a Ziploc bag and stored at 4°C until pYb1s-ndmDCEA completes induction). Finally, the growth curves were examined under an enzyme marker to select the strains that could grow well and produce PX.

Week 3 (6.15-6.21): Collation and analysis of data from screening results

(i) Collation of screening results data

1. Liquid-phase detection

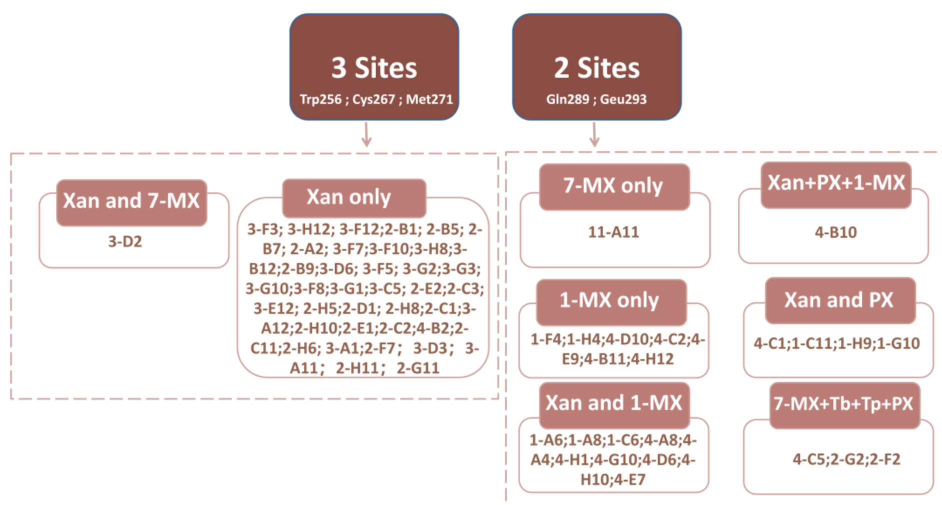
Well-grown strains were screened in 96-well plates and samples were prepared for liquid phase assay. Afterwards, these strains will also be analysed by sequencing to confirm their mutation status and potential functional properties.

Plate 1 (2 sites) / Δ OD_{600} in 24 h

Plate 4 (2 sites) / Δ OD_{600} in 24 h

Plate 2 (3 sites) / Δ OD_{600} in 24 h

Plate 3 (3 sites) / Δ OD_{600} in 24 h



2. Sequencing analysis

Several mutants were randomly selected for sequencing under each set of products.

Mutant Name	Mutain Site1:	Mutain Site 2:	Mutain Site 3:	Mutain Site 4:	Mutain Site 5:	Xan Yield	1-MX Yield	7-MX Yield	Tb Yield	PX Yield
1-F4	\	\	\	Gln289→Gly289	Leu293→Pro293	0	0.02	0	0	0
3-G3	Trp256→Thr256	Cys267→Pro267	Met271→Arg271	\	\	0.54	0	0	0	0
3-F12	Trp256→Gly256	Cys267→Leu267	Met271→Phe271	\	\	0.62	0	0	0	0
1-H4	\	\	\	Gln289→Gly289	Leu293→Pro293	0.01	0.01	0	0	0
2-B9	Trp256→Ser256	Cys267→Ala267	Met271→Gly271	\	\	0.5	0	0	0	0
3-D2	Trp256→Thr256	Cys267→Val267	Met271→Leu271	\	\	0.43	0	0	0	0
4-G10	\	\	\	Gln289→Ser289	Leu293→Phe293	0.02	0.02	0	0	0
4-B10	\	\	\	Gln298→Val289	Leu293→Gln293	0.02		0	0	0.02
4-E7	\	\	\	Gln289→Val289	\	0.02	0.02	0	0	0
2-H11	Trp256→Ser256	Met271→Ile271	\	\	\	0.65	0	0	0	0
4-C5	\	\	\	Gln289→Ser289	\	0	0	0.05	0.06	0.77
2-D6	\	\	\	Gln289→Gly289	\	0	0	0.15	0.01	0.19
4-C1	\	\	\	Gln289→Val289	Leu293→Gln293	0.01	0	0	0	0.04
1-C11	\	\	\	Gln289→Met289	Leu293→Arg293	0	0	0	0	0.15
1-H9	\	\	\	Gln289→Leu289	Leu293→Gly293	0	0	0	0	0.02
1-G10	\	\	\	Gln289→Leu289	\	0	0	0	0	0.01
2-G2	\	\	\	Gln289→Ser289	\	0	0	0.23	0.01	0.01
2-F2	\	\	\	Gln289→Ser289	\	0	0	0.01	0.02	0.31
1-A11	\	\	\	Gln289→Cys289		0	0	0.17	0	0.01

Beyond our expectation, the products produced by the strains with higher OD600 values were not only paraxanthines, but also contained other compounds of the methylxanthine family (e.g. 1-MX, 7-MX, etc.). They can also be converted to produce xanthines in the sensing module and thus survive. In order to explore the relationship between protein mutations and the generation of multiple products, and thus to find mutation sites that are more favourable for paraxanthine production, we plan to use the Ridge regression model for machine learning.

Week4 (6.22-6.28): Machine learning to identify key loci

(i) Collation of screening results data

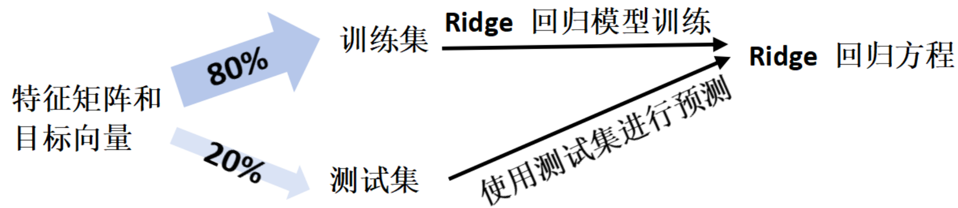
1. Data processing

A large amount of experimental data on protein mutants was first collected and collated, and mutation types were represented by one-hot encoding (one-hot encoding) to generate a feature matrix.

Mutant Name	Mutain Site1：Trp256	Mutain Site 2：Cys267	Mutain Site 3：Met271	Mutain Site 4：Gln289	Mutain Site 5：Leu293	Xan Yield（mM）	1-MX Yield（mM）	7-MX Yield（mM）	Tb Yield（mM）	PX Yield（mM）
1-F4	0	0	0	1	1	0	0.02	0	0	0
3-G3	1	1	1	0	0	0.54	0	0	0	0
3-F12	1	1	1	0	0	0.62	0	0	0	0
1-H4	0	0	0	1	1	0.01	0.01	0	0	0
2-B9	1	1	1	0	0	0.5	0	0	0	0
3-D2	1	1	1	0	0	0.43	0	0	0	0
4-G10	0	0	0	1	1	0.02	0.02	0	0	0
4-B10	0	0	0	1	1	0.02		0	0	0.02
4-E7	0	0	0	1	0	0.02	0.02	0	0	0
2-H11	1	1	0	0	0	0.65	0	0	0	0
4-C5	0	0	0	1	0	0	0	0.05	0.06	0.77
2-D6	0	0	0	1	0	0	0	0.15	0.01	0.19
4-C1	0	0	0	1	1	0.01	0	0	0	0.04
1-C11	0	0	0	1	1	0	0	0	0	0.15
1-H9	0	0	0	1	1	0	0	0	0	0.02
1-G10	0	0	0	1	0	0	0	0	0	0.01
2-G2	0	0	0	1	0	0	0	0.23	0.01	0.01
2-F2	0	0	0	1	0	0	0	0.01	0.02	0.31
1-A11	0	0	0	1	0	0	0	0.17	0	0.01

2. Model Selection and Training

The dataset is randomly divided into a training set and a test set, where the training set is used for model fitting and the test set is used to validate the generalisation performance of the model.



3. Model evaluation

MSE is used as an evaluation metric to measure the predictive accuracy and error distribution of the

- **Xan Yield MSE (Ridge)**: 0.0047
- **1-MX Yield MSE (Ridge)**: 0.0001
- **7-MX Yield MSE (Ridge)**: 0.0010
- **Tb Yield MSE (Ridge)**: 0.0000079
- **PX Yield MSE (Ridge)**: 0.0020

- **Xan产量 R^2 (Ridge)**: 0.941
- **1-MX产量 R^2 (Ridge)**: 0.188
- **7-MX产量 R^2 (Ridge)**: 0.560
- **Tb产量 R^2 (Ridge)**: 0.853
- **PX产量 R^2 (Ridge)**: 0.847

4. Analysis and projection of results