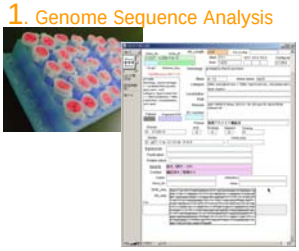
1. Genome Sequence Analysis

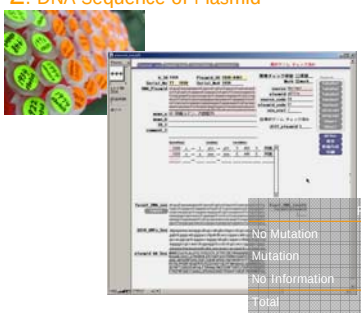


• *Thermus thermophilus* HB8 has about 2200 ORFs.

⇩

• About 2200 Records from the Genome Sequence Analysis

2. DNA Sequence of Plasmid

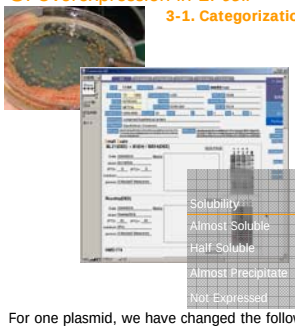


	Records	ORFs
No Mutation	1478	1428
Mutation	459	282
No Information	599	295
Total	2536	1985

• One or more plasmids are constructed from each ORF.
• The relationship between "ORF" and "plasmid sequence" is one-to-many.
• According to records on July 21, 2536 plasmids from 1985 ORFs were constructed, and 1428 ORFs without mutation have been constructed.

3. Overexpression in *E. coli*

3-1. Categorization of Expression Level

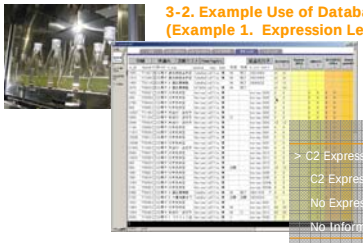


Solubility \ Expression Amount	Large	Small
Almost Soluble	A1(7)	A2(5)
Half Soluble	B1(6)	B2(4)
Almost Precipitate	C1(3)	C2(2)
Not Expressed	D(1)	

For one plasmid, we have changed the following conditions;
- Inducing or Repressing Reagent (IPTG, Lactose, Glucose)
- *E. coli* Strains
- Composition of the Medium or Synthetic Media
- Cultivation Temperature (37℃, 20℃)
- Add tRNAs for Rare Codon

The level of expression was categorized by the amount and solubility of the expressed protein observed by SDS-PAGE.

3-2. Example Use of Database (Example 1. Expression Level)



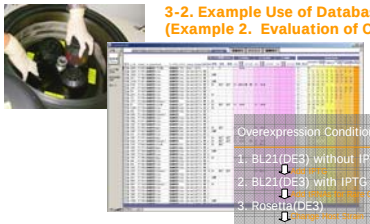
	Records	ORFs
> C2 Expressed	988	930
C2 Expressed	51	47
No Expressed	385	309
No Information	610	475
Total	2034	1761

• Translating the Category into the Numerical Value

• For Example ⇩
"< 3" Represents "Small Expression and Precipitate" or "Not Expressed"

⇩
• Enables the Trial of Next Condition

3-2. Example Use of Database (Example 2. Evaluation of Conditions)

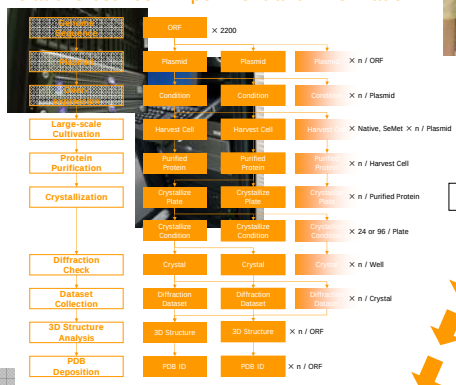


Overexpression Condition	Expressed /Total	Success Rate	(Soluble)
1. BL21(DE3) without IPTG	867/1227	70.7%	(53.9%)
2. BL21(DE3) with IPTG	+8	0.7%	(0.2%)
3. Rosetta(DE3)	+61	5.0%	(3.2%)
4. HMS174(DE3)	+7	0.6%	(0.2%)
5. BL21(DE3)+Glucose	+3	0.2%	(0.2%)
Summary	946/1227	77.1%	(57.7%)

• The priority will be changed if we found a condition which has more high success rate.

START

0. Relations between Experiment and Information

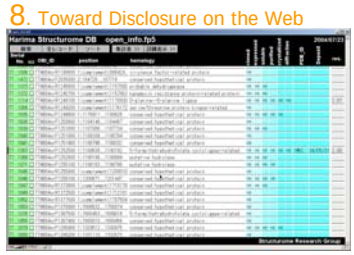


This Database system was build to manage and share the information about
- Genome
- Overexpression
- Purification
- Crystallization
- Large-scale Cultivation
- X-ray Diffraction Datasets from Beamline

for total 2200 ORFs.

GOAL

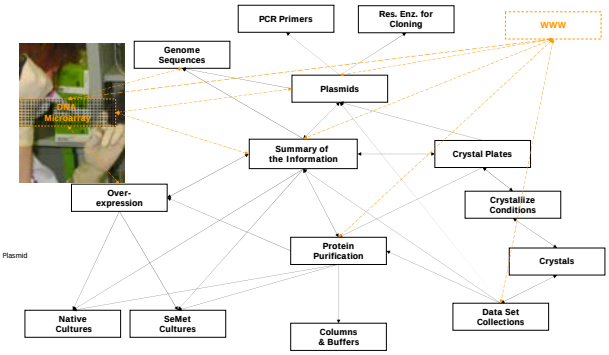
8. Toward Disclosure on the Web



In the future
- The result from the DNA microarray analysis will be included this database system.
- These data will be releesed on website

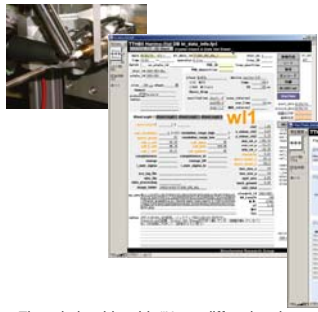
@ http://www.themus.org

7. The Advantage of Constructing DB system by ourselves



• Constructing a database system by ourselves, we can immediately respond to the change of the project situation.

6. Beamline Data Collection



Diffraction Check : 697 Records
Dataset : 107 Records

• The relationship with "X-ray diffraction data sets" and "crystal" is one-to-many.
• Amino acid sequence, crystallization condition, purification process, and the history of measurement for every crystallization plate can be displayed.

5. Crystallization of the Protein

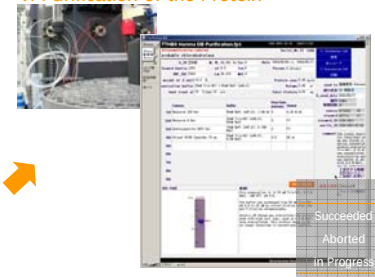
653 Purified Protein Lots / 133 ORFs

1708 Crystallization Plates / 125 ORFs

• The relationship between "purified protein solution" and "crystallization plate" is one-to-many.
• The relationship between "crystallization plate" and "crystallization condition" is one-to-many.

• The days from the crystallized date were calculated, and it prevented forgetting to observe.

4. Purification of the Protein



	Records	ORFs
Succeeded	893	492
Aborted	130	46
In Progress	7	3
Total	1030	541

• The relationship between "harvest cell" and "purification" is one-to-many.
• According to records on July 21, purification of 1030 from 541 ORFs was carried out.
• Purification of 492 ORF has been succeeded.
• It's success rate was 91%.