

博士論文

機械学習を用いたフラビン依存性酸化酵素  
の機能改変

本論文は静岡県立大学大学院薬食生命科学総合学府  
の博士論文である。

2024年3月

静岡県立大学大学院薬食生命科学総合学府  
博士課程 薬学専攻

松下 拓磨

Functional modification of flavin-containing  
monooxygenase through machine learning  
methodology

March, 2024

Takuma Matsushita

## 略語及び語句の定義

BERT : Bidirectional Encoder Representations from Transformers  
BLOSUM : BLOcks SUBstitution Matrix  
CCK-8 : Cell Counting Kit 8  
DMF : *N, N*-dimethyl formamide  
DNA : Deoxyribonucleic acid  
EDTA : ethylenediaminetetraacetic acid  
FAD : Flavin adenine dinucleotide  
FASGAI : Factor Analysis Scales of Generalized Amino acid Information  
FMO : Flavin containing monooxygenase  
GFP : Green Fluorescent Protein  
HPLC : High Performance Liquid Chromatography  
IPTG : Isopropyl- $\beta$ -D-thiogalactopyranoside  
KPB : Potassium phosphate buffer  
MD : Molecular Docking  
MS-WHIM : Molecular Surfaces-Weighted Holistic Invariant Molecular  
NADPH : Nicotinamide adenine dinucleotide phosphate  
Nested-CV : Nested cross validation  
PCR : Polymerase Chain Reaction  
PEG : Polyethylene Glycol  
PF : Photon Factory  
PHBH : *p*-hydroxybenzoate hydroxylase  
ProtFP : Protein FingerPrint  
PSSM : Position-Specific Scoring Matrix  
SDS-PAGE : Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis  
SLS : Swiss Light Source  
ST-scale : Structural Topology scale  
TAPE transformer : Tasks Assessing Protein Embeddings Transformer  
Tris : tris (hydroxymethyl) aminomethane  
t-SNE : t-distributed Stochastic Neighbor Embedding  
VHSE : principal components score Vectors of Hydrophobic, Steric, and  
Electronic properties  
YFP : Yellow Fluorescent Protein

### 使用した培地の組成

LB broth : 10 g/L Trypton, 5 g/L Yeast Extract, 1 g/L NaCl (pH 7.2)

LB/Amp : LB broth containing 100 mg/ml of ampicillin

### 使用した試薬の組成

N3 buffer : 4.2 M Gu-HCl (pH 4.8), 0.9 M potassium acetate

P1 buffer : 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNase A

P2 buffer : 200 mM NaOH, 1% SDS

PB buffer : 5 M Gu-HCl containing 30% isopropanol

PE buffer : 10 mM Tris-HCl (pH 7.5) containing 80% ethanol

TE buffer : 10 mM Tris-HCl (pH 8.0), 1 mM EDTA

### 使用した菌株のゲノタイプ

*E. coli* DH5α : F<sup>-</sup>, Φ80d*lacZ* ΔM15, Δ(*lacZYA-argF*)U169, *deoR*, *recA1*, *endA1*, *hsdR17*(r<sub>K</sub><sup>-</sup>, m<sub>K</sub><sup>+</sup>), *phoA*, *supE44*, λ<sup>-</sup>, *thi-1*, *gyrA96*, *relA1*

*E. coli* BL21(DE3) : F<sup>-</sup>, *ompT*, *hsdS<sub>B</sub>* (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>), *gal* (λcl 857, *ind1*, *Sam7*, *nin5*, *lacUV5-T7gene1*), *dcm* (DE3)

## 目次

### 第一章 序論

第一節	緒言	1
第二節	機械学習を用いた蛍光タンパク質の機能改変	3
第三節	機械学習を用いた多様な配列を持つ変異体酵素の創出	4
第四節	機械学習を用いた p450 の基質特異性の改変	6
第五節	天然物生合成におけるフラビン依存性酸化酵素 (FMO)	8
第六節	引用文献	10

### 第二章 教師データの収集

第一節	FMO のモデル酵素としての PHBH モデル酵素の紹介	13
第二節	PHBH による活性試験の手順	14
第三節	基質誘導体の検討 1	15
第四節	基質誘導体の検討 2	17
第五節	初期のライブラリーの作製	18
第六節	小括	22
第七節	実験項	23
	2.1 PCR による DNA の増幅及び形質転換	23
	2.2 pKW18756 のミニプレップによる精製	23
	2.3 pKW18756-WT の制限酵素処理	24
	2.4 カラムによるプラスミドの精製	24
	2.5 変異型 PHBH のプラスミド作製	24
	2.6 大腸菌を用いた形質転換、培養手順	25
	2.7 PHBH の精製手順	25
	2.8 PHBH の活性試験の手順	26
	<b>Table 2-7-1.</b> pKW18756 作製用のプライマーリスト	27
	<b>Table 2-7-2.</b> 変異体作製用のプライマーリスト	27
	<b>Table 2-7-3.</b> 発現用ベクターと SDS-PAGE のリスト	32
	<b>Figure 2-7-3.</b> 初期の学習用ライブラリー作製のベクターマップ	37
	<b>Figure 2-7-4.</b> 初期ライブラリーの変異体の SDS-PAGE	55
第八節	引用文献	58

### 第三章 機械学習による高活性変異体の探索

第一節	機械学習の概要	59
第二節	Protein descriptor による特徴ベクトルの作成	60
第三節	自然言語処理用深層学習プログラム：BERT による特徴ベクトルの抽出	61
第四節	Nested-cross validation による精度評価	64
第五節	機械学習のモデル関数の構築	66
第六節	モデル関数を用いた全変異体の酵素活性の予想	67
第七節	機械学習により予測された高活性変異体の活性試験による検証	69
第八節	PHBH-MWNL の速度論的な解析	71
第九節	小括	73
第十節	実験項	74
3.1	変異型 PHBH のプラスミド作製	74
3.2	大腸菌の形質転換、培養手順	74
3.3	PHBH の精製手順	74
3.4	PHBH の活性試験の手順	74
3.5	Wild-type PHBH 及び PHBH-MWNL の速度論的な解析	74
	<b>Table 3-10-1.</b> 発現用ベクターと SDS-PAGE のリスト	75
	<b>Figure 3-10-1.</b> セカンドライブラリー作製のベクターマップ	77
	<b>Figure 3-10-2.</b> セカンドライブラリーの SDS-PAGE	82
第十一節	引用文献	83

### 第四章 PHBH-MWNL の活性向上メカニズムの解明

第一節	PHBH-MWNL の結晶化	85
第二節	PHBH-MWNL の共結晶化条件の検討	86
第三節	PHBH-MWNL の共結晶構造の解析	87
第四節	小括	89
第五節	実験項	90
4.1	大腸菌の形質転換、培養手順	90
4.2	PHBH-MWNL の精製手順	90
4.3	結晶化スクリーニング	91
4.4	X 線回折実験及びデータ解析	91
	<b>Figure 4-5-1.</b> SDS-PAGE による PHBH-MWNL の分析	92
	<b>Figure 4-5-2.</b> ゲルろ過クロマトグラフィー	92

<b>Figure 4-5-3.</b> 獲得できた PHBH-MWNL の結晶.....	93
<b>Table 4-5-1.</b> シッティングドロップによる結晶化条件の検討.....	94
<b>Table 4-5-2.</b> ハンギングドロップ蒸気拡散法による結晶化条件の検討 1 ...	94
<b>Table 4-5-3.</b> ハンギングドロップ蒸気拡散法による結晶化条件の検討 2 ...	94
<b>Table 4-5-4.</b> PF BL17A による X 線回折強度の測定条件 .....	95
<b>Table 4-5-5.</b> PHBH-MWNL-6 複合体の結晶学的データと構造精密化の統計値 .....	96
第六節 引用文献.....	97

## 第五章 総括

第一節 本研究の総括.....	99
第二節 今後の展望.....	100
第三節 引用文献.....	101

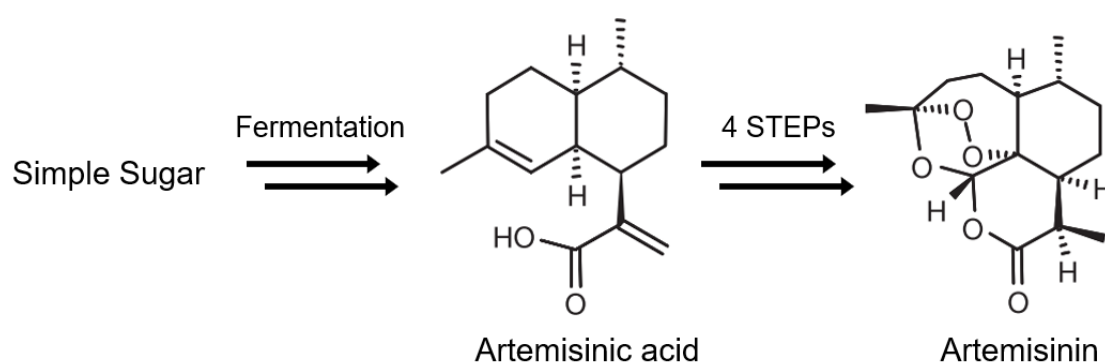
謝辞

# 第一章 序論

## 第一節 緒言

天然物は古くから医薬品のリード化合物として重宝されており、現在上市されている医薬品のうち **50%以上**は天然物由来であることが知られている。しかし、天然物は複雑な骨格や不斉中心を多数持つものが多く、それらの構造を全合成するためには多段階の反応が必要となる。それに対して、生体内では酵素による位置・立体選択的な化学反応により、天然物が生合成されている。さらには、有害な物質の放出が非常に少なく、環境への負担が少ないことから、持続可能な開発の観点からも非常に価値がある。以上の背景から、天然物誘導体の効率的な取得を目指して天然物の生合成研究が行われてきた。

医薬品として承認されている天然物の中には、植物由来の天然成分を含むものも存在するが、生産植物の供給が不安定であり医薬品の安定的な供給の妨げになっているものも存在する。例えば、クソニンジンから単離された **Artemisinin** はマラリアの治療薬<sup>1)</sup>として注目されている。しかし、その生産植物であるクソニンジンの生産量が気候の影響により不安定であることが、医薬品市場への供給の妨げになることが懸念されていた。また、有機合成による手法では、**Artemisinin** の複雑な構造を効率的に合成することが困難であった。そこで、**Jay Keasling** らのグループにより、遺伝子組み換え酵母による発酵法を用いることで、重要中間体である **Artemisinic acid** の生合成が達成された。さらに、**Artemisinic acid** に対して僅か4工程の有機化学的な修飾反応を施すことにより、**Artemisinin** を合成する半合成的手法 (**Figure 1-1-1**)<sup>2)</sup>を用いることで、その問題を解決することに成功した。



**Figure 1-1-1.** 発酵法による Artemisinin の半合成経路

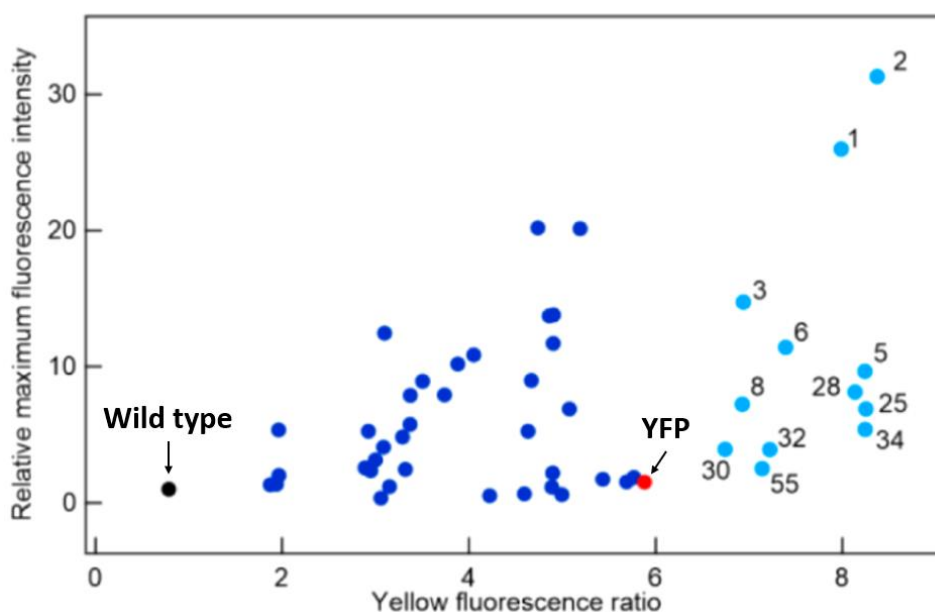


以上のように天然物の生合成酵素は産業的に重要であり、医薬品の効率的な創出を目指して天然物の生合成研究が行われてきた。特に、天然物由来の医薬品は、そのほとんどが本来の天然物に対して化学的な修飾を施した天然物誘導体である。酵素の機能改変により、その基質特異性を変化させることができれば、天然物誘導体の効率的な創出に繋がることが期待される。しかし、二次代謝酵素は基質特異性が非常に高いため、本来の基質とは異なった構造を持つ基質誘導体に対しては、酵素活性が著しく低下することが多い。そこで従来酵素の機能改変では、酵素と基質との相互作用を解明するために、共結晶構造解析が行われてきた。共結晶構造から、基質との相互作用に関与するアミノ酸残基を確認し、各残基の飽和変異体を作製することで、酵素の基質認識性を変化させることが可能であった。しかし、そのような手法では、複数箇所のアミノ酸残基の変異を評価することが難しく、効率的な多重変異体を作製することが困難であった。また、多重変異体を作製する場合には膨大なパターンが存在し、例えば、基質周辺の4アミノ酸残基を改変するだけでも  $20^4=160,000$  の配列が存在する。それに対して研究室レベルで検証できる変異体の数は僅か100-1,000種類程度であり、その数は非常に限られている。以上の背景から、非常に限られた数の変異体を教師データとして用いることで、膨大な配列パターンを網羅的に解析する手法である機械学習が注目されている。

## 第二節 機械学習を用いた蛍光タンパク質の機能改変

機械学習を用いたタンパク質の機能改変の例として、近年産総研の齋藤らによって行われた GFP の機能改変<sup>3)</sup>が挙げられる。この機械学習による GFP の機能改変では、先行研究で明らかにされた GFP の蛍光活性に重要である4つのアミノ酸残基 (S65, S72, H77, T203) に絞って、一重飽和変異体 (19x4=76種類) と多重ランダム変異体 (78種類) を作製した。すると、飽和変異体やランダム変異体ではタンパク質が不安定になって活性試験ができないものも多く存在し、精製できた変異体 GFP についても野生型 GFP よりも蛍光活性が強いものは存在しなかった。

齋藤らはこの蛍光活性を教師データとして機械学習を行い、16万種類の変異体全ての蛍光活性を予測し、予測値が上位の78種類の変異体の中から65種類の変異体の蛍光活性を測定した。すると、65種類全てが野生型よりも強い黄色蛍光を発し、そのうち12種類は YFP<sup>4)</sup>よりも黄色蛍光活性が強いという結果が得られている (Figure 1-2-1)。以前は YFP の作製に膨大な数の GFP の変異体が必要だったのに対して、機械学習を用いればポジティブデータが存在しない状態であっても、効率よく高活性変異体を探索できる可能性が示された。

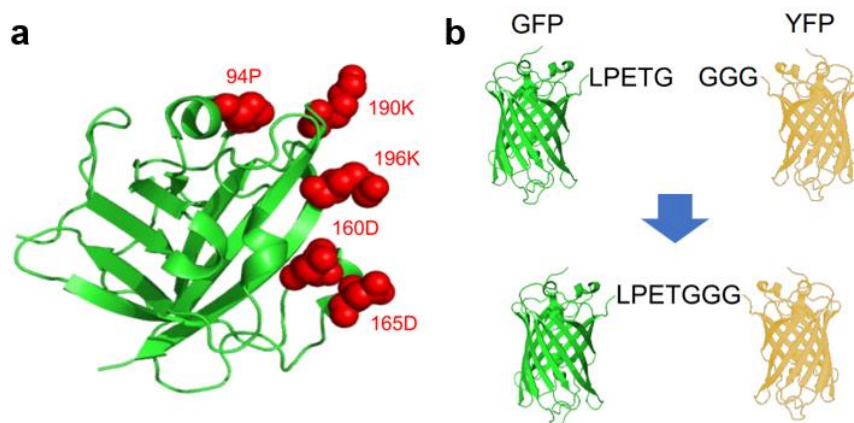


**Figure 1-2-1.** 機械学習により予想された GFP の高活性変異体群  
黒：野生型、赤：YFP、青と水色：機械学習により予想された高活性変異体群  
(YFP よりも黄色蛍光が強い変異体は水色で示している)

### 第三節 機械学習を用いた多様な配列を持つ変異体酵素の創出

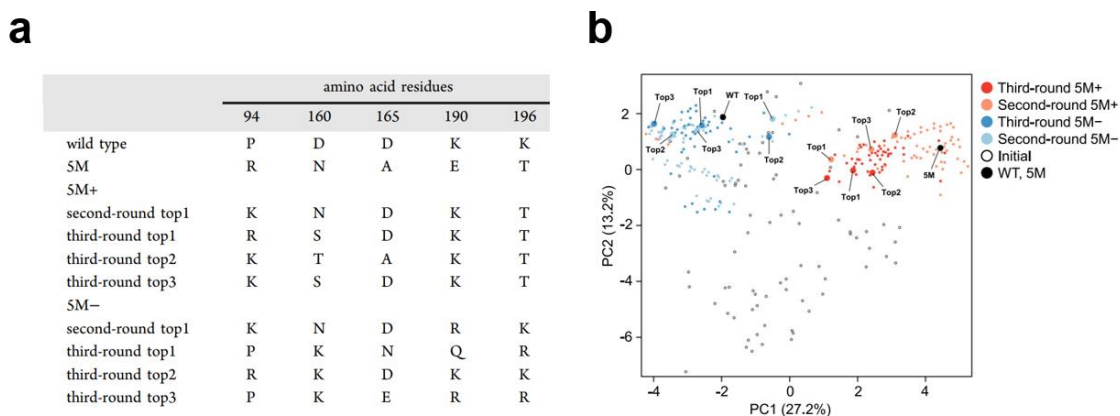
一般的な酵素の機能改変には、結晶構造をもとにして基質認識に関与するアミノ酸残基を特定し、それらの残基を中心にして変異を加えることで高活性変異体を探索する方法やランダムミューテーションで探索する方法がある。前者では見つかった高活性変異体を中心にして、さらに変異を加えていく手法が一般的である。つまり、既存の高活性変異体をベースにして、配列を少しずつ変えていく手法が一般的であるため、既存の高活性変異体と配列が大きく異なる高活性変異体を創出することは困難である。一方、後者は膨大な変異体ライブラリーをスクリーニングする必要があるため、多大な労力が必要であり、また高速なスクリーニング系を樹立できるものに限られる。機械学習を用いた方法ではこれらの制限を受けずに高活性変異体を探索できる。

グラム陽性菌の膜上に存在する **transpeptidase** である **Sortase A (SrtA)** は **LPXTG** 配列を認識して末端の **G** を **GGG** 配列に置き換える作用 (**Figure 1-3-1**) を持ち、タンパク質の修飾やポリマー形成などの産業的な応用が期待されている。しかし、野生型 **SrtA** の酵素活性が弱く、産業的な応用の妨げになっていた。これを改良するためにランダムミューテーションを用いた **SrtA** の機能改変が行われ、 $10^8$  程度の膨大な変異体ライブラリーをスクリーニングすることで、野生型 **SrtA** の 2.5 倍程度の活性を持つ **5M (P94R, D160N, D165A, K190E, K196T)**<sup>5)</sup> が発見された。しかし、**5M** の発見には膨大な労力が必要であり、より高活性なものを得るためにはより効率的な機能改変が必要であることが予想された。また、**5M** を起点にして更なる高活性変異体を探索する方法も考えられたが「局所解」に陥る可能性も考えられた。一方、2021年に報告された産総研の齋藤、東北大の及川らの研究では、機械学習を用いた **SrtA** の機能改変<sup>6)</sup> に成功し、既存の高活性変異体とは異なった変異パターンを持つ高活性変異体を発見することに成功している。



**Figure 1-3-1. (a) SrtA の結晶構造 (b) SrtA によるペプチド鎖の連結反応。**

齋藤らの研究では、既存の高活性変異体が「存在している場合」と「存在していない場合」においてそれぞれ機械学習を行った。活性に重要なアミノ酸残基 (P94, D160, D165, K190, K196) について、一重変異体とランダム変異体を合計 81 種類作製し、機械学習をする際の教師データとして上記の 81 種類に「5M を加えた場合」と「加えなかった場合」において、独立して機械学習 (second round) を行った。機械学習による予測活性をもとにして、それぞれ 67 種類、76 種類を選択し、再度独立して機械学習 (third round) した。両方のモデルにて、機械学習の予測値が上位 50 の変異体をそれぞれ活性試験にて検証した。すると、5M 有で機械学習を行ったライブラリーからは 5M に近い配列の SrtA-RSDKT、5M 無しで機械学習を行ったライブラリーからは 5M と全く異なる SrtA-PKNQR がそれぞれ発見され (Figure 1-3-2)、両者は全く異なる配列を持つが、既存の高活性変異体である 5M の 2 倍以上の酵素活性を持つことが判明した。このことから、機械学習を用いると既存の高活性変異体をさらに改良できるだけでなく、未知の配列を持った高活性変異体の創出も可能であることが示された。



**Figure 1-3-2. (a)** 機械学習により発見された SrtA の高活性変異体 **(b)** PCA による SrtA の高活性変異体のマッピング

## 第四節 機械学習を用いた P450 の基質特異性の改変

近年、日本国内の天然物有機化学を専門とする研究グループでも、機械学習を取り入れる動きが見られている。例えば、東大の小川らの研究<sup>7)</sup>では、緑膿菌由来の toluene や *m*-xylene の代謝経路に着目した。その代謝経路の初速段階は、P450 である XylIM とそのレドックスパートナーである XylIA による酸化反応であり、これが代謝経路の律速段階であると言われている。この XylIM は次段階の反応を触媒する XylIBC に比べると基質認識性が厳格であるため、XylIM の機能改変に成功することがトルエン代謝経路の改変において非常に重要であると考えられている。そこで、機械学習による XylIM の機能改変をすることで、野生型の XylIM がほとんど活性を示さない基質誘導体である 2,6-xyleneol に特異的な変異体酵素の創出に着手した。

小川らの研究では迅速に教師データを収集するために XylIS 及び *mcherry* を用いたレポーターアッセイ (Figure 1-4-1b) を用いた。XylIS は XylIMABC による生成物である 3-methylsalicylic acid を認識することで二量体を形成し、*P<sub>m</sub>* プロモーター領域に結合することで転写を促進<sup>8)</sup>することが報告されている。これを利用して *P<sub>m</sub>* プロモーターの制御下に *mcherry* をおくことで、*mcherry* の蛍光強度から XylIM の活性を評価できるアッセイ系となっている。

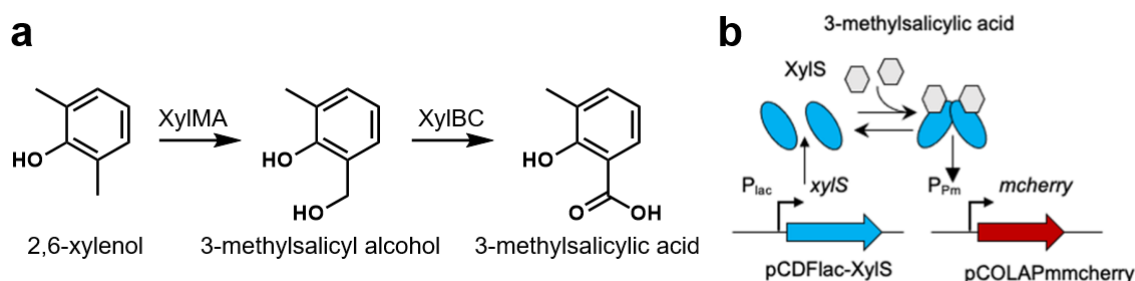
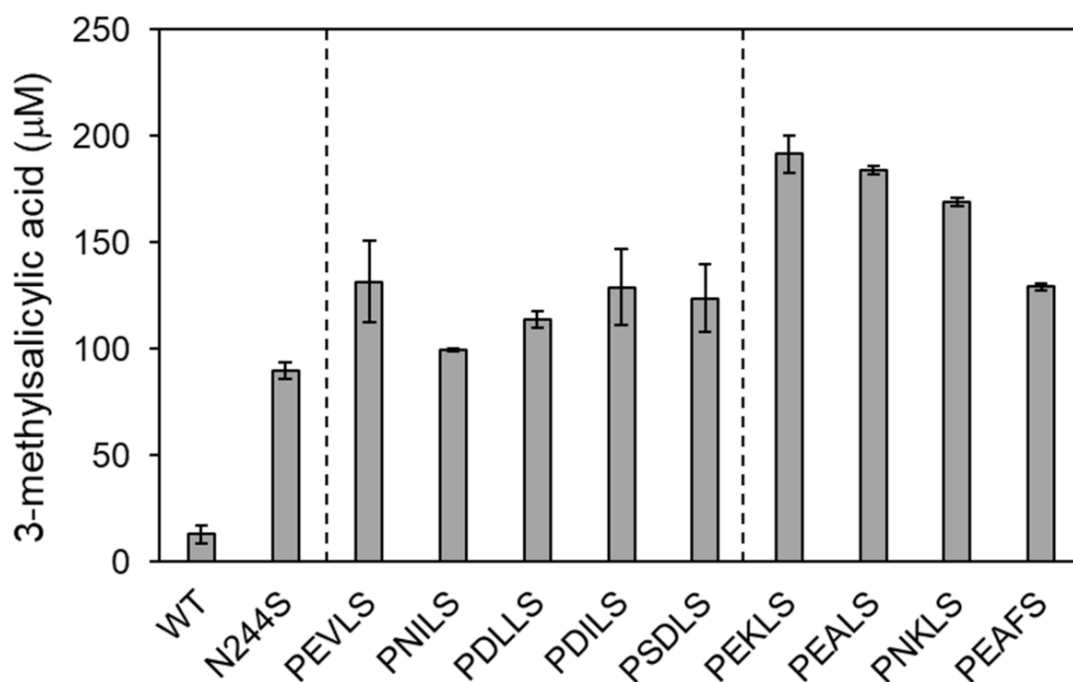


Figure 1-4-1. (a) XylMABC による多段階の酸化反応 (b) XylIS と *mcherry* を組み合わせた 3-methylsalicylic acid の検出過程

小川らの研究では教師データの収集にも工夫が見られる。変異体酵素を作製する場合は、基質結合部位のアミノ酸残基に変異を加えるのが通常の手順であるが、XylIM は結晶構造が解明されておらず、基質との相互作用に関与するアミノ酸残基が不明である。そこで、Phyre 2<sup>9)</sup>を用いて、stearoyl-CoA desaturase<sup>10)</sup>をもとにして、XylIM の立体構造を予測し、活性部位を構成するアミノ酸残基を 10 種類程度まで絞り込んだ。さらに、それらに対してアラニンスキャニングをすることで、酵素活性に重要な 5 つのアミノ酸残基 (P137, D140, V144, F243, N244) を特定した。

野生型酵素に加えて、上記の5アミノ酸残基についての飽和変異体 (19x5=95種類)、ランダム変異体 (31種類) の合計 126種類の変異体酵素を作製し、上記のレポーターアッセイをした。レポーターアッセイの結果を教師データとして機械学習をすることで、5アミノ酸残基の変異体 ( $20^5 = 3,200,000$ 種類) 全ての酵素活性を予測し、その中の上位50種類をレポーターアッセイにて検証した。次に、更なる高活性変異体創出のために、初期ライブラリー (127種類) と機械学習ライブラリー (50種類) の合計 177種類の XyIM を教師データとして用いることで、再度、機械学習を行った。同様に、上位50変異体を選択し、レポーターアッセイにて検証することにより、基質誘導体である 2,6-xyleneol に対する酵素活性が大幅に上昇した XyIM の変異体群を発見することに成功した。

最終的に、レポーターアッセイの結果が良かった変異体に絞って、HPLCにより XyIMABC による最終生成物である 3-methylsalicylic acid を定量することで、野生型 XyIM よりも高活性な変異体 XyIM を多数発見することに成功した (Figure 1-4-2)。このことから、結晶構造が不明な状態であっても機械学習による酵素の機能改変が可能であることが示唆された。



**Figure 1-4-2.** HPLC を用いた XyIM 変異体の活性試験

(左から、初期ライブラリー、1回目の機械学習の変異体、2回目の機械学習の変異体)

## 第五節 天然物生合成におけるフラビン依存性酸化酵素 (FMO)

第二節から第四節で示したように、機械学習を用いることによりタンパク質の機能を効率的に改変させることができ、天然物誘導体の効率的な創出に繋がることが期待される。そこで、天然物の生合成酵素の中でも巨大な酵素ファミリーを形成するフラビン依存性酸化酵素 (FMO) に着目した。FMO は、活性部位に補酵素として FAD や FMN を有しており、それらがもう 1 つの補酵素である NAD(P)H によって還元的に活性化され、分子状酸素が付加することで、活性種であるフラビン過酸化物が形成する。そして、基質から過酸化物への求電子付加反応により、基質に酸素原子が導入される (Figure 1-5-1)。

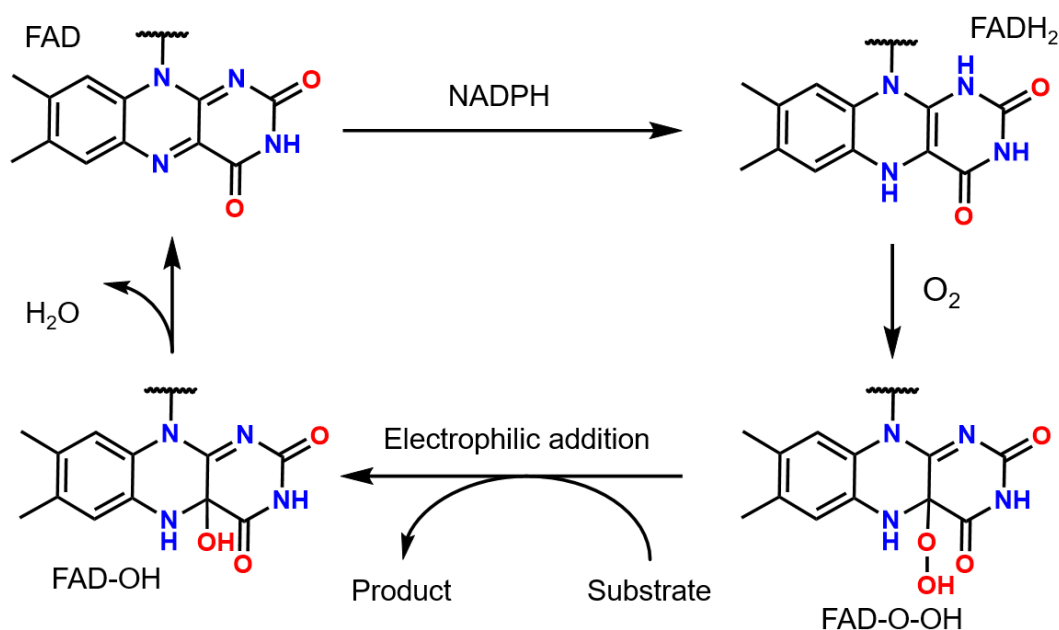
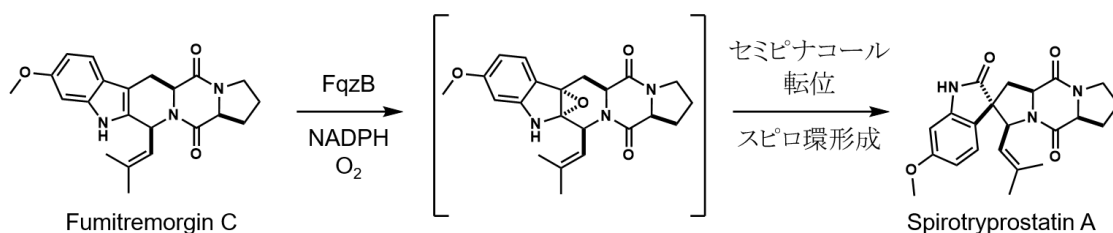


Figure 1-5-1. FMO による求電子付加反応の推定メカニズム

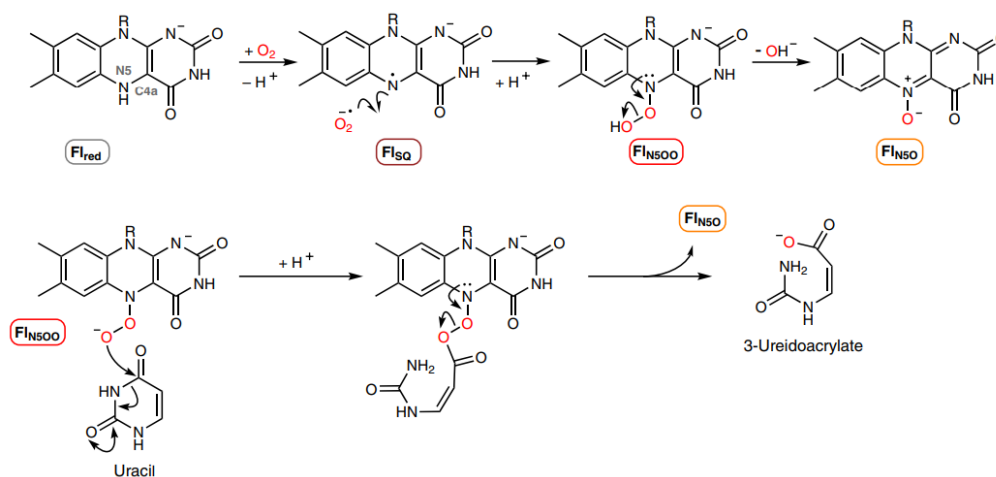
その酵素反応は非常に多岐にわたり、水酸化反応、二重結合の形成、エポキシ環の形成、バイヤービリガー酸化など多種多様な酸化反応を触媒することにより、天然物の構造多様化に大きく貢献している<sup>11)</sup>。さらには、FMO の中には特異な基質認識性を有し、複数の基質をそれぞれの生成物に変換する酵素も報告されている。例えば、*Aspergillus fumigatus* の FqzB<sup>12, 13)</sup>のように、他の生合成クラスター由来の二次代謝産物を基質として認識し、インドール環の二重結合の酸化反応とそれに伴う転位反応により、スピロ環を形成するものも存在する (Figure 1-5-2)。





**Figure 1-5-2.** FqzB による fumitremorgin C の変換反応

通常、FMO は上記の反応機構で示した求電子付加反応により酸化反応を触媒するが、近年、EncM<sup>14, 15</sup>、Rut A<sup>16, 17</sup>など、求核付加反応を触媒するものも報告されている。その反応メカニズムは、FMO の触媒サイクルにおいて、FAD が FADH<sub>2</sub> に変換された後、窒素原子に酸素分子が付加することで flavin N5 oxide が形成され、それに伴う求核付加反応が進行すると推測されている (Figure 1-5-3)。



**Figure 1-5-3.** RutA による Uracil の求核的な加水分解反応

以上から FMO の効率的な機能改変は天然物誘導体の効率的な創出に繋がることが期待される。しかし、FMO の機能改変は難易度が高いことが知られている。先行研究により、FMO は酵素反応時に NADPH による FAD の還元的な活性化をするための構造 (開いた構造) と、それにより形成したフラビンの活性種を酵素の内部に保持する構造 (閉じた構造) を取る必要がある<sup>18, 19</sup>。その際に基質結合部位の一時的な変化を伴うことも多く、結晶構造解析やドッキングシミュレーションなどを用いても詳細な基質認識性の解明は困難であった。機械学習であればこの問題を解決して FMO の効率的な機能改変が達成できると考えられたが、これまでに機械学習を用いた FMO の機能改変は例がない。そこで FMO のモデル酵素を対象として機械学習を用いた機能改変を目指した。



## 第六節 引用文献

- (1) Wang, J.; Xu, C.; Wong, Y. K.; Li, Y.; Liao, F.; Jiang, T.; Tu, Y. Artemisinin, the Magic Drug Discovered from Traditional Chinese Medicine. *Engineering* **2019**, *5*, 32-39.
- (2) Ro, D. K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *5*, 940-943.
- (3) Saito, Y.; Oikawa, M.; Nakazawa, H.; Niide, T.; Kameda, T.; Tsuda, K.; Umetsu, M. Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins. *ACS Synth Biol* **2018**, *7*, 2014-2022.
- (4) Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **2002**, *20*, 87-90.
- (5) Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11399-11404.
- (6) Saito, Y.; Oikawa, M.; Sato, T.; Nakazawa, H.; Ito, T.; Kameda, T.; Tsuda, K.; Umetsu, M. Machine-Learning-Guided Library Design Cycle for Directed Evolution of Enzymes: The Effects of Training Data Composition on Sequence Space Exploration. *ACS Catalysis* **2021**, *11*, 14615-14624.
- (7) Ogawa, Y.; Saito, Y.; Yamaguchi, H.; Katsuyama, Y.; Ohnishi, Y. Engineering the Substrate Specificity of Toluene Degrading Enzyme XylM Using Biosensor XylS and Machine Learning. *ACS Synth. Biol.* **2023**, *12*, 572-582.
- (8) Xue, H.; Shi, H.; Yu, Z.; He, S.; Liu, S.; Hou, Y.; Pan, X.; Wang, H.; Zheng, P.; Cui, C.; Viets, H.; Liang, J.; Zhang, Y.; Chen, S.; Zhang, H. M.; Ouyang, Q. Design, Construction, and Characterization of a Set of Biosensors for Aromatic Compounds. *ACS Synth. Biol.* **2014**, *3*, 1011-1014.

- (9) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols*. **2015**, *10*, 845-858.
- (10) Bai, Y.; McCoy, J. G.; Levin, E. J.; Sobrado, P.; Rajashankar, K. R.; Fox, B. G.; Zhou, M. X-ray structure of a mammalian stearyl-CoA desaturase. *Nature* **2015**, *524*, 252–256.
- (11) Paul, C. E.; Eggerichs, D.; Westphal, A. H.; Tischler, D.; Berkel, W. J. H. Flavoprotein monooxygenases: Versatile biocatalysts. *Biotechnol. Adv.* **2021**, *51*, 107712
- (12) Tsunematsu, Y.; Ishikawa, N.; Wakana, D.; Goda, Y.; Noguchi, H.; Moriya, H.; Hotta, K.; Watanabe, K. Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.* **2013**, *9*, 818–825.
- (13) Matsushita, T.; Kishimoto, S.; Hara, K.; Hashimoto, H.; Watanabe, K. Structural and functional analyses of a spiro-carbon-forming, highly promiscuous epoxidase from fungal natural product biosynthesis. *Biochemistry* **2020**, *59*, 4787-4792.
- (14) Teufel, R.; Miyanaga, A.; Michaudel, Q.; Stull, F.; Louie, G.; Noel, J. P.; Baran, P. S.; Palfey, B.; Moore, B. S. Flavin-mediated dual oxidation controls an enzymatic Favorskii-type rearrangement. *Nature* **2013**, *503*, 552–556.
- (15) Teufel, R.; Stull, F.; Meehan, M. J.; Michaudel, Q.; Dorrestein, P. C.; Palfey, B.; Moore, B. S. Biochemical establishment and characterization of EncM's flavin-N5-oxide cofactor. *J. Am. Chem. Soc.* **2015**, *137*, 8078–8085.
- (16) Mukherjee, T.; Zhang, Y.; Abdelwahed, S.; Ealick, S. E.; Begley, T. P. Catalysis of a Flavoenzyme-Mediated Amide Hydrolysis. *J. Am. Chem. Soc.* **2010**, *132*, 5550–5551.
- (17) Matthews, A.; Saleem-Batcha, R.; Sanders, J. N.; Stull, F.; Houk, K. N.; Teufel, R. Aminoperoxide adducts expand the catalytic repertoire of flavin

monooxygenases. *Nat. Chem. Biol.* **2020**, *16*, 556–563.

(18) Wang, J.; Ortiz-Maldonado, M.; Entsch, B.; Massey, V.; Ballou, D.; Gatti, D. L. Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 608-613.

(19) Gatti, D. L.; Palfey, B. A.; Lah, M. S.; Entsch, B.; Massey, V.; Ballou, D. P.; Ludwig, M. L. The mobile flavin of 4-OH benzoate hydroxylase. *Science* **1994**, *266*, 110-114.

## 第二章 教師データの収集

### 第一節 FMO のモデル酵素としての PHBH

本研究では機械学習を用いた酵素の機能改変を達成すべく、FMO のモデル酵素である緑膿菌由来の PHBH<sup>1)</sup>に着目した (Figure 2-1-1a)。この PHBH は FMO のモデル酵素として広く研究されていて、様々な知見が集まっていることから、今回の研究に適していることが予想された。PHBH は基質である 4-ヒドロキシ安息香酸のメタ位に水酸基を導入して 3,4-ジヒドロキシ安息香酸に変換する (Figure 2-1-1b)。よって、野生型酵素がほとんど反応しないような基質誘導体を発見し、次にその基質誘導体に対して高活性な変異体酵素を創出できれば、酵素の機能改変に成功したことになる。まずは、野生型酵素の基質特異性について確認すべく、基質誘導体を用いた活性試験に着手した。

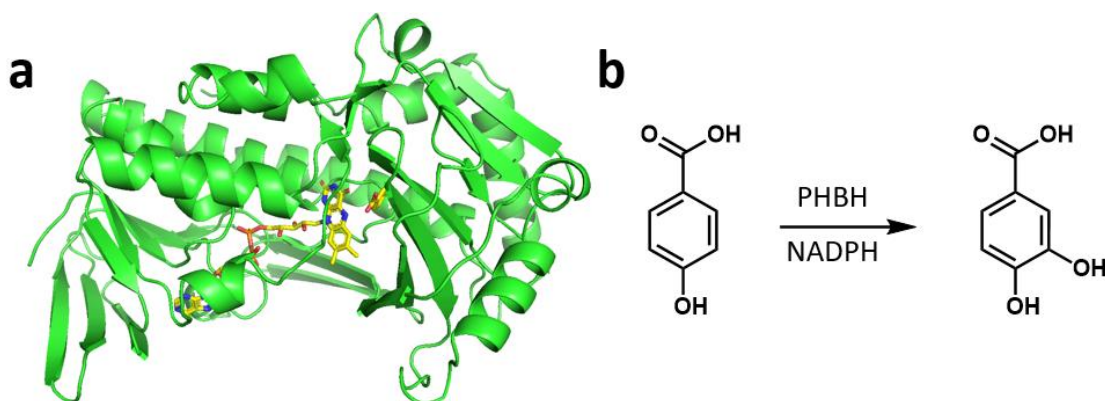


Figure 2-1-1. (a) PHBH の全体構造。 (b) PHBH による 4-ヒドロキシ安息香酸から 3,4-ジヒドロキシ安息香酸への変換反応。

## 第二節 PHBH による活性試験の手順

まずは、野生型 PHBH の基質特異性について確認すべく、活性試験に用いる基質誘導体を検討することにした。FMO の酵素活性を定量する簡易的な方法としては、NADPH が持つ 340 nm の紫外線吸収を定量することで、酵素反応の前後で消費された NADPH を定量する方法が良く行われている。しかし、FMO の触媒サイクルでは、アンカップリング<sup>2)</sup>の経路が存在し、その経路では生成物の代わりに過酸化水素が放出されてしまう。それにより、NADPH の消費量と生成物の量が必ずしも一致せず、NADPH の消費量をもとにして生成物を正確に定量することが困難である。また、アンカップリングを定量するための手法として、西洋ワサビペルオキシダーゼを用いて、アンカップリングにより形成された過酸化水素を定量する方法<sup>3)</sup>も存在する。しかし、近年の研究ではアンカップリングにより過酸化水素だけでなく、スーパーオキシドが形成する<sup>2)</sup>ということも報告されており、紫外線吸収による FMO の活性試験の定量を困難にしている。

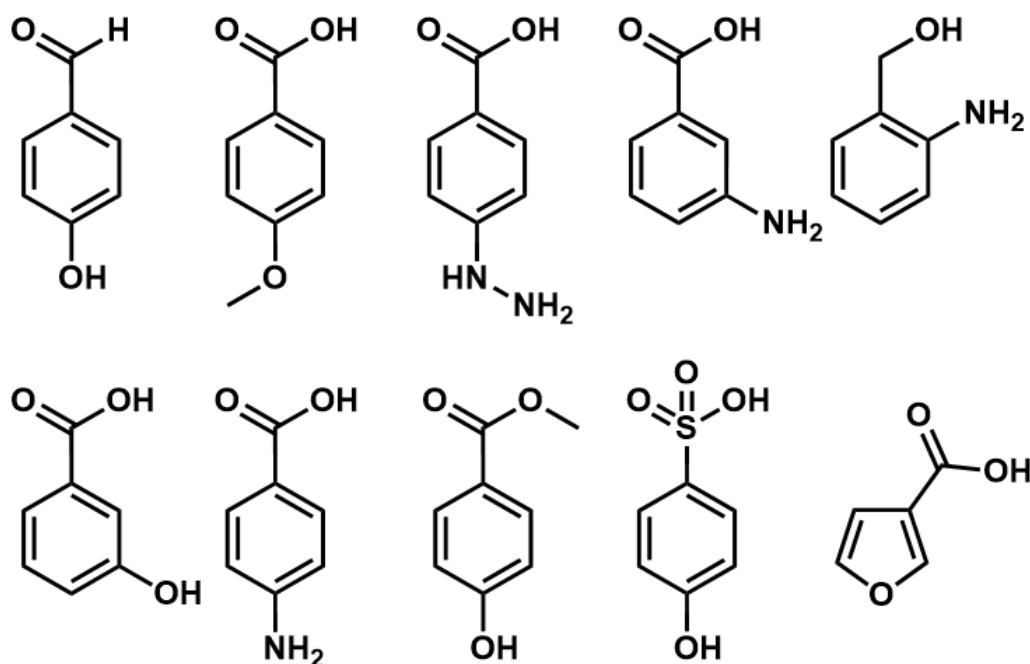
以上の背景から、詳細な解析をすべく NADPH の消費量と生成物の収量の両方を定量することにした。活性試験が終了後、反応溶液を 2 つに分割し、一方は HPLC による生成物の定量に用い、もう一方は NADPH の消費量の定量に用いた。NADPH の消費量は上述のように 340 nm の紫外線吸収により定量するのが一般的である。しかし、今回は生成物が 340 nm の紫外線吸収を持つため、NADPH の紫外線吸収と重なってしまい、正確な定量が困難であることが予想される。そこで、CCK-8 を用いて残存する NADPH をホルマザン色素に変換し (Figure 2-2-1)、生成したホルマザン色素の 450 nm の可視光線の吸収を測定することで、残存する NADPH の量を定量した。



Figure 2-2-1. CCK-8 による NADPH のホルマザン色素への変換反応

### 第三節 基質誘導体の検討 1

野生型 PHBH の基質認識性について確認するために、入手可能な基質誘導体を用いた活性試験に着手した。基質誘導体としては、基質のカルボキシ基がアルデヒド基、スルホ基、メチルエステルに置換された化合物や、水酸基がメトキシ基、アミノ基に置換されたもの、および水酸基の位置がメタ位になった化合物を用いた (**Figure 2-3-1**)。これらの基質誘導体を用いて活性試験に着手したが、NADPH の消費は一切確認することができなかった。

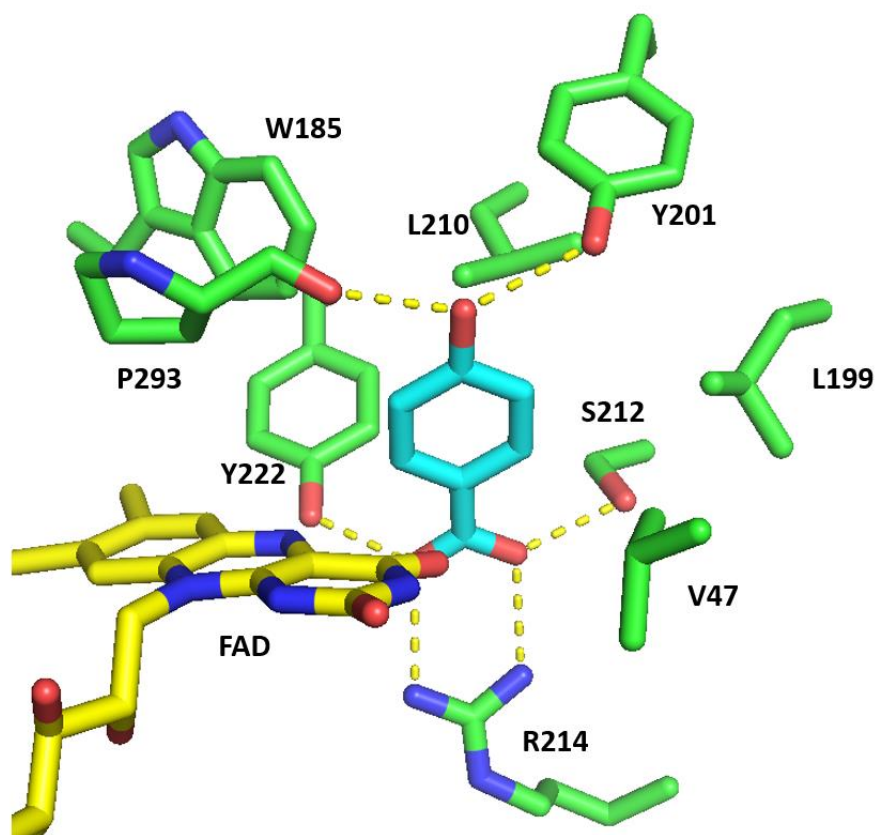


**Figure 2-3-1.** 野生型酵素と反応しなかった基質誘導体の構造

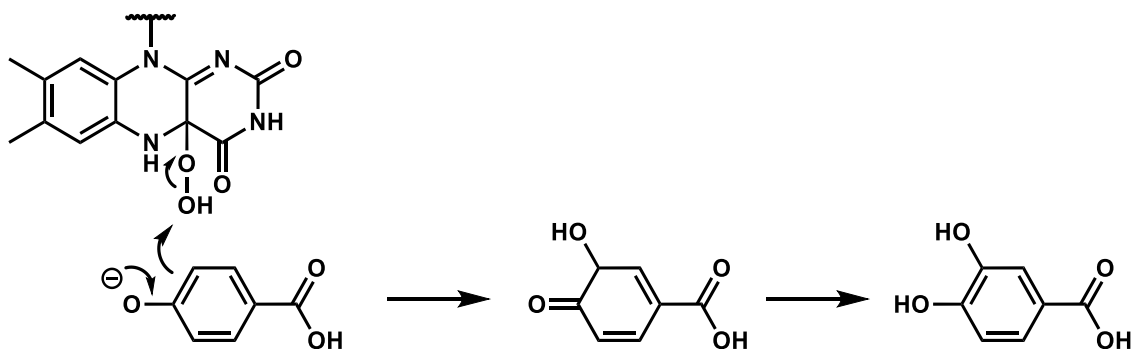
先行研究では、PHBH の基質結合部位では、基質のカルボキシ基の周辺で、水素結合を形成する極性アミノ酸残基 (S212, R214, Y222) を確認でき (**Figure 2-3-2**)<sup>1)</sup>、これらが基質との相互作用に関与することが推測される。よって、基質のカルボキシ基が別の官能基に置換されることにより、基質を活性部位に保持することができなくなったと思われる。

また、基質の水酸基は、Y201 の側鎖や P293 の主鎖との水素結合を形成することが確認できた (**Figure 2-3-2**)。先行研究では、プロトンリレーによる水酸基の脱プロトン化と、それに伴う水酸化反応 (**Figure 2-3-3**) により酵素反応が進行する。よって、水酸基の位置が変わることにより、酵素反応が進行しなくなったと予想される。また、水酸基がアミノ基に置換された基質誘導体では酵素反応が進行しなかった。これは、水酸基よりもアミノ基の方が酸解離定数が

小さいため、脱プロトン化を受けにくいことが原因であると考えられる。



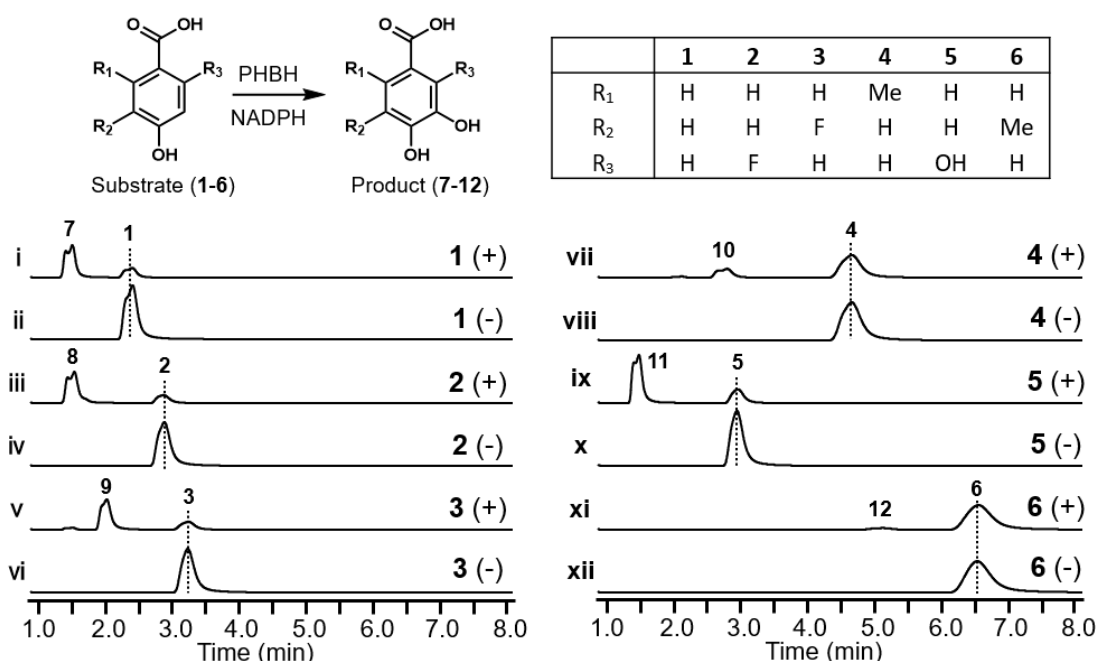
**Figure 2-3-2.** 野生型 PHBH の基質結合部位 (アミノ酸残基は緑、補酵素 FAD は黄色、基質は青)



**Figure 2-3-3.** 野生型 PHBH による 4-hydroxybenzoic acid の水酸化反応

## 第四節 基質誘導体の検討2

第三節で示したように、カルボキシ基と4位の水酸基が酵素の基質認識に必要であることが示唆されたので、本来の基質である4-ヒドロキシ安息香酸 (**1**)の構造を基本骨格として、基質誘導体を検討することにした。**1**のベンゼン環上の水素原子がフッ素、メチル基、水酸基などに置換された基質誘導体 (**1-6**)を用いて活性試験をし、それぞれの生成物 (**7-12**)を定量した。その結果、野生型酵素は基質誘導体 **1-5** に対しては良好な活性を示したが、**6** に対しては活性が弱い (**Figure 2-4-1**) ことが確認できた。よって、野生型の酵素活性が弱い **6** に対する高活性変異体の創出を目指して機械学習に着手した。



**Figure 2-4-1.** 野生型 PHBH と基質誘導体の活性試験であり、生成物量は 260 nm の紫外線吸収にて定量した。ネガティブコントロールとして、熱失活した酵素を用いている。

(i) 野生型 PHBH と **1**, (ii) 熱失活した野生型 PHBH と **1**, (iii) 野生型 PHBH と **2**, (iv) 熱失活した野生型 PHBH と **2**, (v) 野生型 PHBH と **3**, (vi) 熱失活した野生型 PHBH と **3**, (vii) 野生型 PHBH と **4**, (viii) 熱失活した野生型 PHBH と **4**, (ix) 野生型 PHBH と **5**, (x) 熱失活した野生型 PHBH と **5**, (xi) 野生型 PHBH と **6**, (xii) 熱失活した野生型 PHBH と **6**.



## 第五節 初期のライブラリーの作製

機械学習用の初期のライブラリーを作製するために、特定のアミノ酸残基を選択し、それらに対して、飽和変異体とランダム変異体の作製を試みた。共結晶構造から、**1**の周辺の極性のアミノ酸残基として Y201, S212, R214, Y222 が、疎水性のアミノ酸残基として V47, W185, L199, L210 がそれぞれ確認できた

(**Figure 2-3-2**)。極性のアミノ酸残基は、基質との水素結合の形成、基質の水酸基の脱プロトン化などにより酵素反応を触媒することが知られており、これらの変異体の作製では活性が消失することが予想された。よって、疎水性残基 (V47, W185, L199, L210) の部位特異的な変異体を作製することにした。

4アミノ酸残基の変異体を作製する場合は、考えられる変異体として  $20^4=160,000$  通りの膨大なパターンが存在し、その中から偏りなく教師データを収集する必要がある。第一章で紹介した齋藤らによる研究では、野生型酵素及び、特定部位の一重飽和変異体、多重ランダム変異体を作製し、活性試験をすることで教師データを取集している。よって、上記の4アミノ酸残基に対して、一重飽和変異体及び多重ランダム変異体を作製することにした。

一重飽和変異体 76 種類のうち 30 種類は不安定で精製することができなかつたため (**Table 2-5-1**)、精製することができた 46 種類に絞って活性試験にて検証を行った。また、多重ランダム変異体については 88 種類のうち 30 種類は精製することができなかつたため (**Table 2-5-1**)、精製できた 58 種類について活性試験を行った。そして、NADPH の消費量と生成物の量を評価した。不安定で精製することができなかつた変異体は活性をゼロと定義した。

**Table 2-5-1.** 初期のライブラリーの変異体リスト

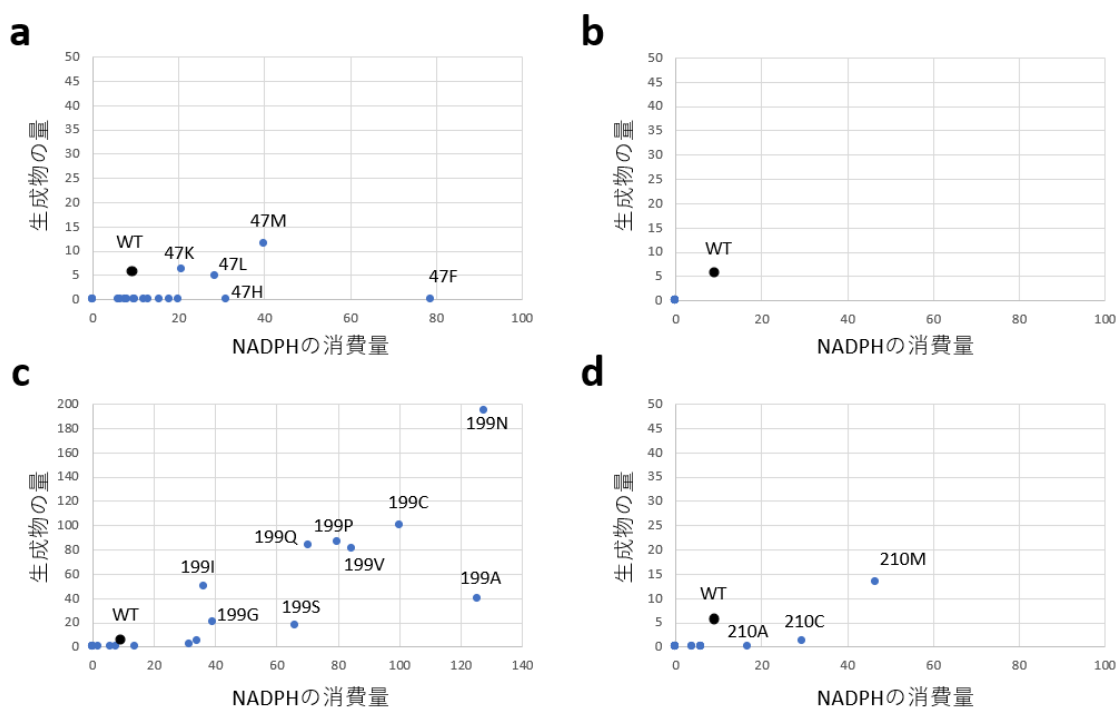
	野生型	一重飽和変異体	ランダム多重変異体	総数
精製できた	1	46	58	105
精製できなかった	0	30	30	60
総数	1	76	88	165

Trp185 の一重変異体の発現量は多かったものの、大部分の変異体が不安定であったため可溶性酵素として得ることができなかつた。唯一 PHBH-W185H は精製することに成功し、活性試験をしたが活性を持っていないことが判明した (**Figure 2-5-1**)。

Val47 や Leu210 の変異体については、野生型酵素よりも活性の上昇につな

がったものが複数確認できた (PHBH-V47K, V47M, L210M など)。また、生成物の量と NADPH の消費量に大きな乖離が存在し、NADPH の消費は野生型酵素よりも増加しているが生成物の生産量は減少している変異体 (PHBH-V47F, V47H, L210C など) が確認できた (Figure 2-5-1)。

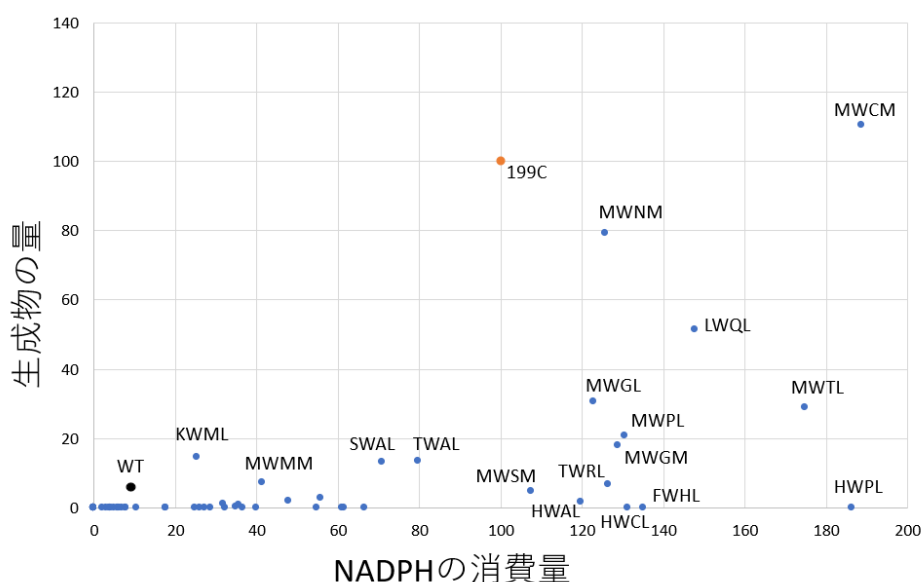
Leu199 の一重変異体では、野生型酵素よりも生成物の量が上昇している変異体を多数確認することができた。また、興味深いことに、L199N や L199Q のような極性のアミノ酸に置換したもので、生成物の量が上昇していた。極性のアミノ酸残基であっても、PHBH-L199D や PHBH-L199E のように酸性のアミノ酸残基に置換されたものでは活性の消失が確認され、そのメカニズムが大変興味深く思われた (Figure 2-5-1)。



**Figure 2-5-1. (a) Val47, (b) Trp185, (c) Leu199, (d) Leu210 の飽和変異体の活性試験の結果 (縦軸が生成物の量、横軸が NADPH の消費量)**

通常の酵素の機能改変では、活性部位の基質周辺の網羅的な変異体を作製し、その中で活性が強かった変異体を起点にして別の個所に変異を導入することで、更なる高活性変異体を創出する手順が用いられる。しかし、一重変異体の活性試験の結果と後述する多重変異体の活性試験の結果を見比べると、予想外の結果になるものが多数存在し、その基質認識性は非常に複雑であることが予想された。

例えば、PHBH-V47M や PHBH-L210M は野生型酵素よりも多くの生成物を生産することが確認できた (**Figure 2-5-2**)。また、初期のライブラリーの一重変異体の中で最も活性が強かったものは PHBH-L199N であった (**Figure 2-5-2**) ので、PHBH-L199N の変異を PHBH-V47M や PHBH-L210M の変異と組み合わせることで、更なる高活性変異体の創出が期待される。しかしながら、PHBH-MWNM (M47, W185, N199, M210) の活性は PHBH-L199N の半分程度であった (**Figure 2-5-2**)。また、PHBH-V47M, PHBH-L199P/S, PHBH-L210M は野生型酵素よりも活性が強かったが、PHBH-MWPM (M47, W185, P199, M210) や PHBH-MWSM (M47, W185, S199, M210) の活性は野生型酵素よりも弱く (**Figure 2-5-2**)、活性が高い変異を組み合わせても、必ずしも活性が上昇しないことが示唆された。それに対して、PHBH-V47T や PHBH-L199R は生成物の収量がほとんどゼロであったが、PHBH-TWRL (T47, W185, R199, L210) は野生型酵素よりも多くの生成物を生産した (**Figure 2-5-2**)。この結果からも多重変異体の基質認識性は非常に複雑であることが示唆された (**Figure 2-5-2**)。



**Figure 2-5-2.** ランダム変異体の活性試験の結果 (縦軸が生成物の量、横軸が NADPH の消費量)

76種類の一重飽和変異体の中で、6に対する活性が野生型酵素よりも高いものは12種類であり、88種類の多重ランダム変異体の中では、野生型酵素よりも活性が高いものは13種類であった。また、初期のライブラリーの結果から、アンカップリングの影響で生成物の量とNADPHの消費量が一致していない変異体が多く、また不一致の大きさも変異体ごとに大きく異なっていることが分かる。例えば、一重変異体では、NADPHの消費量と生成物の収量の間には緩やかな正の相関が確認できるのに対して、多重変異体の場合はその相関が弱くなっており、野生型酵素よりも多くのNADPHが消費されているにもかかわらず、生成物が確認できないような変異体も複数確認されている。例えば、PHBH-HWPLのNADPHの消費量は初期のライブラリーの中で最も多いにもかかわらず、生成物の量はゼロであった（Figure 2-5-2）。別の言い方をすると、多重変異体の方がアンカップリングが発生しやすい傾向にあり、ランダムな多重変異体の作製では、アンカップリングが少なく高活性の変異体酵素を創出することが困難であることが予想された。

## 第六節 小括

機械学習用のデータ収集をするために、基質誘導体の活性試験に着手をした。本来の基質と官能基の位置や種類が違う基質誘導体においては、酵素活性を確認することができなかつたので、**1**の構造を基本骨格として基質誘導体の探索をした。基質誘導体 (**1-6**) を用いて野生型酵素との活性試験をし、それぞれの生成物 (**7-12**) を定量すると、**3**位にメチル基を有する基質誘導体 **6** に対する酵素活性が弱いことが判明した。そのため、**6** に対する高活性変異体の創出を目指して、**76** 種類の飽和変異体と **88** 種類のランダム変異体の創出をした。**6** に対する変換効率が野生型よりも高い変異体酵素を創出することに成功したが、その割合は低く、**164** 種類中 **24** 種類であった。よって、従来の方法では効率的な酵素の機能改変が困難であることが予想された。

## 第七節 実験項

### 2.1 PCR による DNA の増幅及び形質転換

*Pseudomonas aeruginosa* のゲノム DNA をテンプレートとして、相補的なプライマーである pKW18756-WT-F と pKW18756-WT-R (Table 2-7-1) を用いて、PCR にて目的の遺伝子配列を増幅した。1.2%のアガロース電気泳動により、目的の位置にバンドを確認できた (Figure 2-7-1)。PCR 溶液と pET-21c の空ベクターを混ぜたもので *E. coli* DH5  $\alpha$  を形質転換し、LB/Amp 平板培地にて選択をした。コロニーを取り、2 mL の LB/Amp 液体培地に加え、37 °C, 180 rpm にてオーバーナイトで前培養した後、下記の手順にてミニプレップをすることにより目的のプラスミド DNA を獲得した。

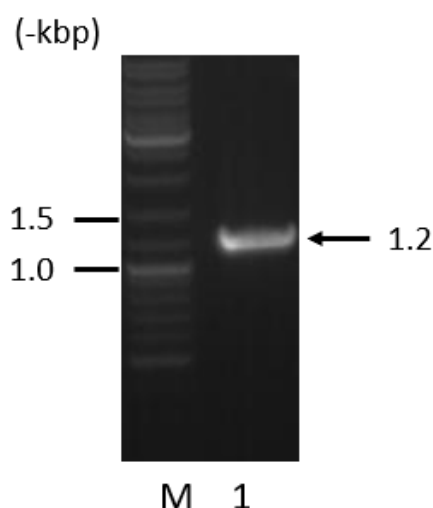


Figure 2-7-1. M : マーカー、1 : PCR 産物

### 2.2 pKW18756 のミニプレップによる精製

形質転換した *E. coli* DH5  $\alpha$  を LB/Amp 平板培地に播種し、37 °C で一晩培養した。生えてきたコロニーを 2 mL の LB/Amp 液体培地に加え、37 °C, 180 rpm にて一晩培養した。2 mL のチューブに大腸菌の菌体を集菌し、P1 buffer を 250  $\mu$ L を加えて、菌体を懸濁した後、P2 buffer を 250  $\mu$ L を加えて十分に転倒混和し、菌体を溶菌した。N3 buffer を 350  $\mu$ L 加えて中和した後、15,000 G, 20 min, 4 °C で遠心し、菌体などの不溶物を沈殿した。上清を取り、等量の冷えたイソプロパノールを加えて、プラスミド DNA を不溶化し、15,000 G, 20 min, 4 °C で遠心して沈殿させた。上清を除いた後、70%エタノールを加えて沈殿をリンスした。上清を取り除き、残ったエタノールを減圧下で濃縮した後、TE buffer を 50  $\mu$ L 加えて、プラスミド DNA を溶解した。

### 2.3 pKW18756-WT の制限酵素処理

0.5  $\mu\text{L}$  の Apa I, 0.5  $\mu\text{L}$  の Pst I, 0.5  $\mu\text{L}$  の Xba I と 1  $\mu\text{L}$  の 10X buffer、3  $\mu\text{L}$  のプラスミド DNA を混和し、超純水にて 10  $\mu\text{L}$  までメスアップした。37  $^{\circ}\text{C}$ , 2 hours 程度プラスミド DNA を消化した後 5  $\mu\text{L}$  を取り、1.2%のアガロースゲルにて電気泳動し、目的の位置にバンドが出ていることを確認した (Figure 2-7-2)。

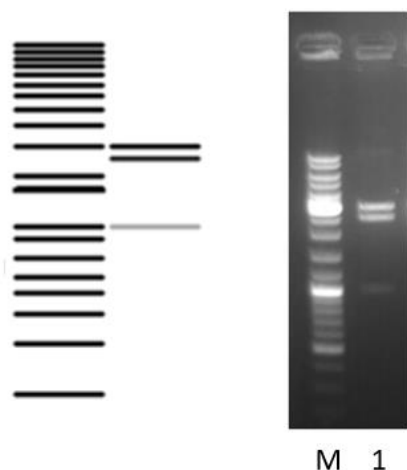


Figure 2-7-2. M : マーカー、1 : PCR 産物

### 2.4 カラムによるプラスミドの精製

ミニプレップにて精製したプラスミド DNA に、3 M の酢酸ナトリウムを 15  $\mu\text{L}$ , PB buffer を 100  $\mu\text{L}$  加えて、プラスミド DNA を不溶化した後、全量をエコノカラムに移した。10,000 rpm, 1 min, 4  $^{\circ}\text{C}$  で遠心し、プラスミド DNA をエコノカラムの担体に吸着させた。100  $\mu\text{L}$  の PB buffer、100  $\mu\text{L}$  の PE buffer の順に、エコノカラムに加えて、同様に 10,000 rpm, 1 min, 4  $^{\circ}\text{C}$  で遠心することにより不純物をリンスした。その後 30  $\mu\text{L}$  の超純水を加えて、同様に 10,000 rpm, 1 min, 4  $^{\circ}\text{C}$  で遠心することにより、プラスミド DNA を溶出した。

### 2.5 変異型 PHBH のプラスミド作製

Premix 5  $\mu\text{L}$ , 2 pmol/L の DNA primer (Forward and Reverse) をそれぞれ 1  $\mu\text{L}$  ずつ、100 分の 1 に希釈したプラスミド溶液を 1  $\mu\text{L}$  を混ぜて超純水にて 10  $\mu\text{L}$  までメスアップした。PCR をすることにより、プラスミド DNA を増幅した。PCR 産物を 1.5  $\mu\text{L}$  取り、1.2%のアガロース電気泳動することにより、DNA の増幅を確認した。PCR 産物 8.5  $\mu\text{L}$  と 10x buffer 1  $\mu\text{L}$ 、DpnI 0.5  $\mu\text{L}$  を混合し、37  $^{\circ}\text{C}$ , 2 hours で鑄型となるプラスミド DNA を消化した。その後、80  $^{\circ}\text{C}$ , 15 min 程度加熱することで Dpn I を失活させ、全量をケミカルコンピン

トセル *E. coli* DH5 $\alpha$  に加えた。氷中で 15 min 静置し、42 °C, 30 sec ヒートショックし、LB/Amp 平板培地に播種した。37 °C で一晩培養し、翌朝、コロニー形成を確認した。

## 2.6 大腸菌を用いた形質転換、培養手順

プラスミド DNA を 0.5  $\mu$ L を取り、*E. coli* BL21(DE3) に加え、1.5 kV でエレクトロポレーションした。これを LB/Amp 平板培地に塗布し、37 °C で一晩培養した。生えてきたコロニーを取り、2 mL の LB/Amp 液体培地に植菌し、37 °C, 180 rpm で一晩振とう培養した。前培養液を 0.5 mL ずつ取り、20 mL の LB/Amp 液体培地に、それぞれ植菌し、37 °C, 180 rpm で振とう培養した。OD<sub>600</sub> が 0.5-0.6 に到達後、十分に氷冷した。終濃度が 0.1 mM になるように IPTG を加え、18 °C, 180 rpm, 20 hours 振とう培養した。その後、培養液を 50 mL のファルコンチューブに移し、2,500 G, 4 °C, 10 min で遠心分離した。上清を捨て、沈殿を懸濁し、2 mL チューブに移して、15,000 G, 4 °C, 1 min で遠心分離した。上清を捨て -80 °C で保管した。

## 2.7 PHBH の精製手順

破砕用緩衝液 (100 mM Tris-HCl pH 7.4 (4 °C), 100 mM NaCl, 50  $\mu$ M FAD) を 800  $\mu$ L 取り、大腸菌を懸濁した。超音波破砕し、10,000 G, 4 °C, 10 min で遠心分離した。上清を別の 1.5 mL チューブに移し、50  $\mu$ L の Ni-NTA agarose を加え、低温室で一時間以上転倒混和した。その後、15,000 G, 4 °C, 1 min で遠心分離し、上清を捨てた。すると、Ni-NTA agarose が黄色に着色していることが確認できた。破砕用緩衝液に溶解した 20 mM imidazole を 500  $\mu$ L 加えて、軽く懸濁した後、15,000 G, 4 °C, 1 min で遠心分離し、上清を捨てた。その後、破砕用緩衝液に溶解した 200 mM imidazole を 200  $\mu$ L 加えて、軽く懸濁した。すると、Ni-NTA agarose が元の色に戻っているため、タンパク質が溶出できたことが確認できた。再度、同様に 10,000 G, 4 °C, 1 min で遠心分離した後、タンパクを含んだ上清を取った。アミコンにタンパク溶液を移し、400  $\mu$ L になるまで破砕用緩衝液を加えた。10,000 G, 4 °C で 100-150  $\mu$ L になるまで遠心した。再度、400  $\mu$ L になるまで破砕用緩衝液を加えて、同様に 10,000 G, 4 °C で 100-150  $\mu$ L になるまで遠心した。これをイミダゾールの終濃度が 1 mM 以下になるまで繰り返した。SDS-PAGE にて精製した変異体酵素の純度を確認した後、ブラッドフォード法にて酵素の濃度を定量した。



## 2.8 PHBH の活性試験の手順

1  $\mu$ M PHBH, 1 mM NADPH, 1 mM **6**, 100 mM KPB (pH 7.4), 100 mM NaCl (100  $\mu$ L) で 25 °C, 30 min 活性試験した。酵素反応終了後 溶液を半分に分け、一方は等量の DMF を加え酵素反応を停止した後、10  $\mu$ L の CCK-8 を加えて、残存する NADPH をホルマザン色素に変換した。90  $\mu$ L を 96 穴プレートに移し、450 nm の紫外線吸収を定量した。

もう一方は、1 N HCl を 5  $\mu$ L 加えて pH を下げた後、100  $\mu$ L の酢酸エチルを用いて 2 回抽出した。抽出後の有機層を減圧下で濃縮した後、50  $\mu$ L の DMF に溶解することで HPLC 用のサンプルを調整した。NADPH の消費量が確認できなかったサンプルは、HPLC による測定をせずに生成物の量をゼロと定義した。NADPH の消費が確認できたサンプルは、ACQUITY UPLC HSS (1.8  $\mu$ m, 2.1  $\times$  50 mm) C18 reversed-phase column を用いて、5% MeCN, 0.05% formic acid の移動相で、0.50 mL/min の流速で 260 nm の紫外線吸収をもとにして、生成物を定量した。

**Table 2-7-1.** pKW18756 作製のプライマーリスト (*phbh* の遺伝子とアニーリングする配列を大文字で表記)

No	Primer name	sequence
1	pKW18756-WT-F	5'-aggagatatacaATGAAGACTCAAGTCGCCATCATCG-3'
2	pKW18756-WT-R	5'-gtgggtggtggtgatgCTCGATTTCTCGTAGGGCAGGC-3'

**Table 2-7-2.** 変異体作製のプライマーリスト (変異が入っている配列を小文字で表記)

No	Primer name	sequence
1	pKW18756-V47A-F	5'-GCaGGaGcaCTGGAACAGGGTATGGTCGACCT-3'
2	pKW18756-V47A-R	5'-TTCCAGtgCtCctGCGCGGATGCGGCCG-3'
3	pKW18756-V47C-F	5'-GCaGGatgtCTGGAACAGGGTATGGTCGACCT-3'
4	pKW18756-V47C-R	5'-TTCCAGacatCctGCGCGGATGCGGCCG-3'
5	pKW18756-V47D-F	5'-GCaGGaGatCTGGAACAGGGTATGGTCGACCT-3'
6	pKW18756-V47D-R	5'-TTCCAGatCtCctGCGCGGATGCGGCCG-3'
7	pKW18756-V47E-F	5'-GCaGGaGaaCTGGAACAGGGTATGGTCGACCT-3'
8	pKW18756-V47E-R	5'-TTCCAGttCtCctGCGCGGATGCGGCCG-3'
9	pKW18756-V47F-F	5'-GCaGGatTtCTGGAACAGGGTATGGTCGACCT-3'
10	pKW18756-V47F-R	5'-TTCCAGaAatCctGCGCGGATGCGGCCG-3'
11	pKW18756-V47G-F	5'-GcTGGtGgaCTGGAACAGGGTATGGTCGACCT-3'
12	pKW18756-V47G-R	5'-TTCCAGtcCaCCaGCGCGGATGCGGCCG-3'
13	pKW18756-V47H-F	5'-GCaGGacatCTGGAACAGGGTATGGTCGACCT-3'
14	pKW18756-V47H-R	5'-TTCCAGatgtCctGCGCGGATGCGGCCG-3'
15	pKW18756-V47I-F	5'-GCaGGaaTtCTGGAACAGGGTATGGTCGACCT-3'
16	pKW18756-V47I-R	5'-TTCCAGaAttCctGCGCGGATGCGGCCG-3'
17	pKW18756-V47K-F	5'-GCaGGaaaaCTGGAACAGGGTATGGTCGACCT-3'
18	pKW18756-V47K-R	5'-TTCCAGttttCctGCGCGGATGCGGCCG-3'
19	pKW18756-V47L-F	5'-GCaGGacTaCTGGAACAGGGTATGGTCGACCT-3'
20	pKW18756-V47L-R	5'-TTCCAGtAgtCctGCGCGGATGCGGCCG-3'
21	pKW18756-V47M-F	5'-GCaGGaaTGCTGGAACAGGGTATGGTCGACCT-3'
22	pKW18756-V47M-R	5'-TTCCAGCAttCctGCGCGGATGCGGCCG-3'
23	pKW18756-V47N-F	5'-GCaGGaaatCTGGAACAGGGTATGGTCGACCT-3'
24	pKW18756-V47N-R	5'-TTCCAGatttCctGCGCGGATGCGGCCG-3'
25	pKW18756-V47P-F	5'-GCaGGtccGCTGGAACAGGGTATGGTCGAC-3'
26	pKW18756-V47P-R	5'-TTCCAGCggaCctGCGCGGATGCGGCCG-3'

27	pKW18756-V47Q-F	5'-GCaGGacaaCTGGAACAGGGTATGGTCGACCT-3'
28	pKW18756-V47Q-R	5'-TTCCAGttgtCctGCGCGGATGCGGCCG-3'
29	pKW18756-V47R-F	5'-GCaGGacgtCTGGAACAGGGTATGGTCGACCT-3'
30	pKW18756-V47R-R	5'-TTCCAGacgtCctGCGCGGATGCGGCCG-3'
31	pKW18756-V47S-F	5'-GCaGGttcGCTGGAACAGGGTATGGTCGACCT-3'
32	pKW18756-V47S-R	5'-TTCCAGCgaaCctGCGCGGATGCGGCCG-3'
33	pKW18756-V47T-F	5'-GCaGGaacGCTGGAACAGGGTATGGTCGACCT-3'
34	pKW18756-V47T-R	5'-TTCCAGCgttCctGCGCGGATGCGGCCG-3'
35	pKW18756-V47W-F	5'-GCaGGatgGCTGGAACAGGGTATGGTCGACCT-3'
36	pKW18756-V47W-R	5'-TTCCAGCcatCctGCGCGGATGCGGCCG-3'
37	pKW18756-V47Y-F	5'-GCaGGatatCTGGAACAGGGTATGGTCGACCT-3'
38	pKW18756-V47Y-R	5'-TTCCAGatatCctGCGCGGATGCGGCCG-3'
39	pKW18756-W185A-F	5'-TTtGGtgcaCTaGGCCTGCTCGCCGACAC-3'
40	pKW18756-W185A-R	5'-GCctAGtgcaCCaAACGGATAGACCCGCTCGAAGACC-3'
41	pKW18756-W185C-F	5'-TTtGGaTGtCTaGGCCTGCTCGCCGACAC-3'
42	pKW18756-W185C-R	5'-GCctAGaCatCCaAACGGATAGACCCGCTCGAAGACC-3'
43	pKW18756-W185D-F	5'-TTtGGagatCTaGGCCTGCTCGCCGACAC-3'
44	pKW18756-W185D-R	5'-GCctAGatctCCaAACGGATAGACCCGCTCGAAGACC-3'
45	pKW18756-W185E-F	5'-TTtGGagaaCTaGGCCTGCTCGCCGACAC-3'
46	pKW18756-W185E-R	5'-GCctAGttctCCaAACGGATAGACCCGCTCGAAGACC-3'
47	pKW18756-W185F-F	5'-TTtGGaTttCTaGGCCTGCTCGCCGACAC-3'
48	pKW18756-W185F-R	5'-GCctAGaaAtCCaAACGGATAGACCCGCTCGAAGACC-3'
49	pKW18756-W185G-F	5'-TTtGGaggaCTaGGCCTGCTCGCCGACAC-3'
50	pKW18756-W185G-R	5'-GCctAGtcctCCaAACGGATAGACCCGCTCGAAGACC-3'
51	pKW18756-W185H-F	5'-TTtGGacatCTaGGCCTGCTCGCCGACAC-3'
52	pKW18756-W185H-R	5'-GCctAGatgtCCaAACGGATAGACCCGCTCGAAGACC-3'
53	pKW18756-W185I-F	5'-TTtGGaattCTaGGCCTGCTCGCCGACAC-3'
54	pKW18756-W185I-R	5'-GCctAGaattCCaAACGGATAGACCCGCTCGAAGACC-3'
55	pKW18756-W185K-F	5'-TTtGGaaaaCTaGGCCTGCTCGCCGACAC-3'
56	pKW18756-W185K-R	5'-GCctAGttttCCaAACGGATAGACCCGCTCGAAGACC-3'
57	pKW18756-W185L-F	5'-TTtGGacttCTaGGCCTGCTCGCCGACAC-3'
58	pKW18756-W185L-R	5'-GCctAGaagtCCaAACGGATAGACCCGCTCGAAGACC-3'
59	pKW18756-W185M-F	5'-TTtGGaatGCTaGGCCTGCTCGCCGACAC-3'
60	pKW18756-W185M-R	5'-GCctAGCattCCaAACGGATAGACCCGCTCGAAGACC-3'
61	pKW18756-W185N-F	5'-TTtGGaatCTaGGCCTGCTCGCCGACAC-3'

62	pKW18756-W185N-R	5'-GCctAGattCCaAACGGATAGACCCGCTCGAAGACC-3'
63	pKW18756-W185P-F	5'-TTtGGtcctCTaGGCCTGCTCGCCGACAC-3'
64	pKW18756-W185P-R	5'-GCctAGaggaCCaAACGGATAGACCCGCTCGAAGACC-3'
65	pKW18756-W185Q-F	5'-TTtGGacaaCTaGGCCTGCTCGCCGACAC-3'
66	pKW18756-W185Q-R	5'-GCctAGttgtCCaAACGGATAGACCCGCTCGAAGACC-3'
67	pKW18756-W185R-F	5'-TTtGGaaGaCTaGGCCTGCTCGCCGACAC-3'
68	pKW18756-W185R-R	5'-GCctAGtCttCCaAACGGATAGACCCGCTCGAAGACC-3'
69	pKW18756-W185S-F	5'-TTtGGaaGtCTaGGCCTGCTCGCCGACAC-3'
70	pKW18756-W185S-R	5'-GCctAGaCttCCaAACGGATAGACCCGCTCGAAGACC-3'
71	pKW18756-W185T-F	5'-TTtGGaactCTaGGCCTGCTCGCCGACAC-3'
72	pKW18756-W185T-R	5'-GCctAGagttCCaAACGGATAGACCCGCTCGAAGACC-3'
73	pKW18756-W185V-F	5'-TTtGGagttCTaGGCCTGCTCGCCGACAC-3'
74	pKW18756-W185V-R	5'-GCctAGaactCCaAACGGATAGACCCGCTCGAAGACC-3'
75	pKW18756-W185Y-F	5'-TTtGGaTatCTaGGCCTGCTCGCCGACAC-3'
76	pKW18756-W185Y-R	5'-GCctAGatAtCCaAACGGATAGACCCGCTCGAAGACC-3'
77	pKW18756-L199A-F	5'-CACGAAgcaATCTACGCCAACCATCCGCGGGCTT-3'
78	pKW18756-L199A-R	5'-GTAGATgctTTCGTGGCTGACCGGCGGGGTGT-3'
79	pKW18756-L199C-F	5'-CACGAAgtATCTACGCCAACCATCCGCGGGCTT-3'
80	pKW18756-L199C-R	5'-GTAGATacaTTCGTGGCTGACCGGCGGGGTGT-3'
81	pKW18756-L199D-F	5'-CACGAAgatATCTACGCCAACCATCCGCGGGCTT-3'
82	pKW18756-L199D-R	5'-GTAGATatcTTCGTGGCTGACCGGCGGGGTGT-3'
83	pKW18756-L199E-F	5'-CACGAAgaaATCTACGCCAACCATCCGCGGGCTT-3'
84	pKW18756-L199E-R	5'-GTAGATtctTTCGTGGCTGACCGGCGGGGTGT-3'
85	pKW18756-L199F-F	5'-CACGAAttATCTACGCCAACCATCCGCGGGCTT-3'
86	pKW18756-L199F-R	5'-GTAGATaAaTTCGTGGCTGACCGGCGGGGTGT-3'
87	pKW18756-L199G-F	5'-CACGAAggaATCTACGCCAACCATCCGCGGGCTT-3'
88	pKW18756-L199G-R	5'-GTAGATtccTTCGTGGCTGACCGGCGGGGTGT-3'
89	pKW18756-L199H-F	5'-CACGAAcAtATCTACGCCAACCATCCGCGGGCTT-3'
90	pKW18756-L199H-R	5'-GTAGATatGTTTCGTGGCTGACCGGCGGGGTGT-3'
91	pKW18756-L199I-F	5'-CACGAAaTaATCTACGCCAACCATCCGCGGGCTT-3'
92	pKW18756-L199I-R	5'-GTAGATtAtTTCGTGGCTGACCGGCGGGGTGT-3'
93	pKW18756-L199K-F	5'-CACGAAaaaATCTACGCCAACCATCCGCGGGCTT-3'
94	pKW18756-L199K-R	5'-GTAGATtttTTCGTGGCTGACCGGCGGGGTGT-3'
95	pKW18756-L199M-F	5'-CACGAAaTGATCTACGCCAACCATCCGCGGGCTT-3'
96	pKW18756-L199M-R	5'-GTAGATCaTTCGTGGCTGACCGGCGGGGTGT-3'

97	pKW18756-L199N-F	5'-CACGAAaatATCTACGCCAACCATCCGCGGGCTT-3'
98	pKW18756-L199N-R	5'-GTAGATattTTCGTGGCTGACCGCGGGGTGT-3'
99	pKW18756-L199P-F	5'-CACGAACcaATCTACGCCAACCATCCGCGGGCTT-3'
100	pKW18756-L199P-R	5'-GTAGATtgTTCGTGGCTGACCGCGGGGTGT-3'
101	pKW18756-L199Q-F	5'-CACGAACaaATCTACGCCAACCATCCGCGGGCTT-3'
102	pKW18756-L199Q-R	5'-GTAGATttTTCGTGGCTGACCGCGGGGTGT-3'
103	pKW18756-L199R-F	5'-CACGAAagaATCTACGCCAACCATCCGCGGGCTT-3'
104	pKW18756-L199R-R	5'-GTAGATtctTTCGTGGCTGACCGCGGGGTGT-3'
105	pKW18756-L199S-F	5'-CACGAAatctATCTACGCCAACCATCCGCGGGCTT-3'
106	pKW18756-L199S-R	5'-GTAGATagaTTCGTGGCTGACCGCGGGGTGT-3'
107	pKW18756-L199T-F	5'-CACGAAactATCTACGCCAACCATCCGCGGGCTT-3'
108	pKW18756-L199T-R	5'-GTAGATagtTTCGTGGCTGACCGCGGGGTGT-3'
109	pKW18756-L199V-F	5'-CACGAAgTGATCTACGCCAACCATCCGCGGGCTT-3'
110	pKW18756-L199V-R	5'-GTAGATCaTTCGTGGCTGACCGCGGGGTGT-3'
111	pKW18756-L199W-F	5'-CACGAAtgGATCTACGCCAACCATCCGCGGGCTT-3'
112	pKW18756-L199W-R	5'-GTAGATCcaTTCGTGGCTGACCGCGGGGTGT-3'
113	pKW18756-L199Y-F	5'-CACGAAtatATCTACGCCAACCATCCGCGGGCTT-3'
114	pKW18756-L199Y-R	5'-GTAGATataTTCGTGGCTGACCGCGGGGTGT-3'
115	pKW18756-L210A-F	5'-TTtGCagctTGCAGCCAGCGTTCGGCGA-3'
116	pKW18756-L210A-R	5'-GCTGCAagctGCaAAGCCGCGGGATGGTTGGC-3'
117	pKW18756-L210C-F	5'-TTtGCatgtTGCAGCCAGCGTTCGGCGA-3'
118	pKW18756-L210C-R	5'-GCTGCAacatGCaAAGCCGCGGGATGGTTGGC-3'
119	pKW18756-L210D-F	5'-TTtGCagatTGCAGCCAGCGTTCGGCGA-3'
120	pKW18756-L210D-R	5'-GCTGCAatctGCaAAGCCGCGGGATGGTTGGC-3'
121	pKW18756-L210E-F	5'-TTtGCagaaTGCAGCCAGCGTTCGGCGA-3'
122	pKW18756-L210E-R	5'-GCTGCAAttctGCaAAGCCGCGGGATGGTTGGC-3'
123	pKW18756-L210F-F	5'-TTtGCatTtTGCAGCCAGCGTTCGGCGA-3'
124	pKW18756-L210F-R	5'-GCTGCAaAatGCaAAGCCGCGGGATGGTTGGC-3'
125	pKW18756-L210G-F	5'-TTtGCaggaTGCAGCCAGCGTTCGGCGA-3'
126	pKW18756-L210G-R	5'-GCTGCAtcctGCaAAGCCGCGGGATGGTTGGC-3'
127	pKW18756-L210H-F	5'-TTtGCaCatTGCAGCCAGCGTTCGGCGA-3'
128	pKW18756-L210H-R	5'-GCTGCAatGtGCaAAGCCGCGGGATGGTTGGC-3'
129	pKW18756-L210I-F	5'-TTtGCaaTaTGCAGCCAGCGTTCGGCG-3'
130	pKW18756-L210I-R	5'-GCTGCAAtAttGCaAAGCCGCGGGATGGTTGG-3'
131	pKW18756-L210K-F	5'-TTtGCaaaaTGCAGCCAGCGTTCGGCGA-3'

132	pKW18756-L210K-R	5'-GCTGCAAtttGCaAAGCCGCGCGGATGGTTGGC-3'
133	pKW18756-L210M-F	5'-TTtGCaaTGTGAGCCAGCGTTCGGGC-3'
134	pKW18756-L210M-R	5'-GCTGCACAttGCaAAGCCGCGCGGATGGTTGGC-3'
135	pKW18756-L210N-F	5'-TTtGCaaatTGCAGCCAGCGTTCGGCGA-3'
136	pKW18756-L210N-R	5'-GCTGCAatttGCaAAGCCGCGCGGATGGTTGGC-3'
137	pKW18756-L210P-F	5'-TTtGCaCcaTGCAGCCAGCGTTCGGCGA-3'
138	pKW18756-L210P-R	5'-GCTGCAtgGtGCaAAGCCGCGCGGATGGTTGGC-3'
139	pKW18756-L210Q-F	5'-TTtGCaCaaTGCAGCCAGCGTTCGGCGA-3'
140	pKW18756-L210Q-R	5'-GCTGCAAttGtGCaAAGCCGCGCGGATGGTTGGC-3'
141	pKW18756-L210R-F	5'-TTtGCaCgaTGCAGCCAGCGTTCGGCGA-3'
142	pKW18756-L210R-R	5'-GCTGCAtcGtGCaAAGCCGCGCGGATGGTTGGC-3'
143	pKW18756-L210S-F	5'-TTtGCatctTGCAGCCAGCGTTCGGCGA-3'
144	pKW18756-L210S-R	5'-GCTGCAagatGCaAAGCCGCGCGGATGGTTGGC-3'
145	pKW18756-L210T-F	5'-TTtGCaaactTGCAGCCAGCGTTCGGCGA-3'
146	pKW18756-L210T-R	5'-GCTGCAagttGCaAAGCCGCGCGGATGGTTGGC-3'
147	pKW18756-L210V-F	5'-TTtGCagTaTgtAGCCAGCGTTCGGCGACCC-3'
148	pKW18756-L210V-R	5'-GCTaCAActGCaAAGCCGCGCGGATGGTTGG-3'
149	pKW18756-L210W-F	5'-TTtGCatgGTGAGCCAGCGTTCGGCG-3'
150	pKW18756-L210W-R	5'-GCTGCACcatGCaAAGCCGCGCGGATGGTTGGC-3'
151	pKW18756-L210Y-F	5'-TTtGCatatTGCAGCCAGCGTTCGGCGA-3'
152	pKW18756-L210Y-R	5'-GCTGCAatatGCaAAGCCGCGCGGATGGTTGGC-3'

**Table 2-7-3. 発現用ベクターと SDS-PAGE のリスト**

No	Vector map	Vector name	SDS-PAGE (Figure number- Gel name and lane number)	Mutation
1	Figure S2-A1	pKW18756-WT	Figure S4-A1	Wild type
2	Figure S2-B1	pKW18756-V47A	Figure S4-B1	V47A
3	Figure S2-B2	pKW18756-V47C	Figure S4-B2	V47C
4	Figure S2-B3	pKW18756-V47D	Unstable	V47D
5	Figure S2-B4	pKW18756-V47E	Figure S4-B3	V47E
6	Figure S2-B5	pKW18756-V47F	Figure S4-B4	V47F
7	Figure S2-B6	pKW18756-V47G	Unstable	V47G
8	Figure S2-B7	pKW18756-V47H	Figure S4-C1	V47H
9	Figure S2-B8	pKW18756-V47I	Figure S4-C2	V47I
10	Figure S2-B9	pKW18756-V47K	Figure S4-C3	V47K
11	Figure S2-B10	pKW18756-V47L	Figure S4-C4	V47L
12	Figure S2-B11	pKW18756-V47M	Figure S4-C5	V47M
13	Figure S2-B12	pKW18756-V47N	Figure S4-D1	V47N
14	Figure S2-B13	pKW18756-V47P	Unstable	V47P
15	Figure S2-B14	pKW18756-V47Q	Figure S4-D2	V47Q
16	Figure S2-B15	pKW18756-V47R	Figure S4-D3	V47R
17	Figure S2-B16	pKW18756-V47S	Figure S4-D4	V47S
18	Figure S2-B17	pKW18756-V47T	Figure S4-E1	V47T
19	Figure S2-B18	pKW18756-V47W	Figure S4-E2	V47W
20	Figure S2-B19	pKW18756-V47Y	Figure S4-E3	V47Y
21	Figure S2-C1	pKW18756-W185A	Unstable	W185A
22	Figure S2-C2	pKW18756-W185C	Unstable	W185C
23	Figure S2-C3	pKW18756-W185D	Unstable	W185D
24	Figure S2-C4	pKW18756-W185E	Unstable	W185E
25	Figure S2-C5	pKW18756-W185F	Unstable	W185F
26	Figure S2-C6	pKW18756-W185G	Unstable	W185G
27	Figure S2-C7	pKW18756-W185H	Figure S4-F1	W185H
28	Figure S2-C8	pKW18756-W185I	Unstable	W185I
29	Figure S2-C9	pKW18756-W185K	Unstable	W185K
30	Figure S2-C10	pKW18756-W185L	Unstable	W185L

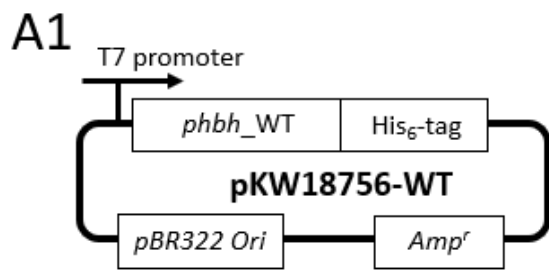
31	Figure S2-C11	pKW18756-W185M	Unstable	W185M
32	Figure S2-C12	pKW18756-W185N	Unstable	W185N
33	Figure S2-C13	pKW18756-W185P	Unstable	W185P
34	Figure S2-C14	pKW18756-W185Q	Unstable	W185Q
35	Figure S2-C15	pKW18756-W185R	Unstable	W185R
36	Figure S2-C16	pKW18756-W185S	Unstable	W185S
37	Figure S2-C17	pKW18756-W185T	Unstable	W185T
38	Figure S2-C18	pKW18756-W185V	Unstable	W185V
39	Figure S2-C19	pKW18756-W185Y	Unstable	W185Y
40	Figure S2-D1	pKW18756-L199A	Figure S4-G1	L199A
41	Figure S2-D2	pKW18756-L199C	Figure S4-G2	L199C
42	Figure S2-D3	pKW18756-L199D	Figure S4-G3	L199D
43	Figure S2-D4	pKW18756-L199E	Figure S4-G4	L199E
44	Figure S2-D5	pKW18756-L199F	Figure S4-G5	L199F
45	Figure S2-D6	pKW18756-L199G	Figure S4-G6	L199G
46	Figure S2-D7	pKW18756-L199H	Figure S4-G7	L199H
47	Figure S2-D8	pKW18756-L199I	Figure S4-H1	L199I
48	Figure S2-D9	pKW18756-L199K	Figure S4-H2	L199K
49	Figure S2-D10	pKW18756-L199M	Figure S4-H3	L199M
50	Figure S2-D11	pKW18756-L199N	Figure S4-H4	L199N
51	Figure S2-D12	pKW18756-L199P	Figure S4-I1	L199P
52	Figure S2-D13	pKW18756-L199Q	Figure S4-I2	L199Q
53	Figure S2-D14	pKW18756-L199R	Figure S4-I3	L199R
54	Figure S2-D15	pKW18756-L199S	Figure S4-I4	L199S
55	Figure S2-D16	pKW18756-L199T	Figure S4-I5	L199T
56	Figure S2-D17	pKW18756-L199V	Figure S4-J1	L199V
57	Figure S2-D18	pKW18756-L199W	Figure S4-J2	L199W
58	Figure S2-D19	pKW18756-L199Y	Figure S4-J3	L199Y
59	Figure S2-E1	pKW18756-L210A	Figure S4-K1	L210A
60	Figure S2-E2	pKW18756-L210C	Figure S4-K2	L210C
61	Figure S2-E3	pKW18756-L210D	Unstable	L210D
62	Figure S2-E4	pKW18756-L210E	Unstable	L210E
63	Figure S2-E5	pKW18756-L210F	Figure S4-K3	L210F
64	Figure S2-E6	pKW18756-L210G	Figure S4-K4	L210G
65	Figure S2-E7	pKW18756-L210H	Figure S4-K5	L210H



66	Figure S2-E8	pKW18756-L210I	Unstable	L210I
67	Figure S2-E9	pKW18756-L210K	Figure S4-K6	L210K
68	Figure S2-E10	pKW18756-L210M	Figure S4-L1	L210M
69	Figure S2-E11	pKW18756-L210N	Unstable	L210N
70	Figure S2-E12	pKW18756-L210P	Unstable	L210P
71	Figure S2-E13	pKW18756-L210Q	Figure S4-L2	L210Q
72	Figure S2-E14	pKW18756-L210R	Unstable	L210R
73	Figure S2-E15	pKW18756-L210S	Figure S4-L3	L210S
74	Figure S2-E16	pKW18756-L210T	Unstable	L210T
75	Figure S2-E17	pKW18756-L210V	Unstable	L210V
76	Figure S2-E18	pKW18756-L210W	Unstable	L210W
77	Figure S2-E19	pKW18756-L210Y	Figure S4-L4	L210Y
78	Figure S2-F1	pKW18756-AAKC	Unstable	V47A/W185A/L199K/L210C
79	Figure S2-F2	pKW18756-AGAC	Unstable	V47A/W185G/L199A/L210C
80	Figure S2-F3	pKW18756-AWSL	Figure S4-M1	V47A/L199S
81	Figure S2-F4	pKW18756-CELF	Unstable	V47C/W185E/L210F
82	Figure S2-F5	pKW18756-CWFL	Figure S4-M2	V47C/L199F
83	Figure S2-F6	pKW18756-CWHL	Figure S4-M3	V47C/L199H
84	Figure S2-F7	pKW18756-CWTL	Figure S4-M4	V47C/L199T
85	Figure S2-F8	pKW18756-DSNE	Unstable	V47D/W185S/L199N/L210E
86	Figure S2-F9	pKW18756-EGGM	Unstable	V47E/W185G/L199G/L210M
87	Figure S2-F10	pKW18756-EKCI	Unstable	V47E/W185K/L199C/L210I
88	Figure S2-F11	pKW18756-EWCL	Figure S4-M5	V47E/L199C
89	Figure S2-F12	pKW18756-EWNL	Figure S4-M6	V47E/L199N
90	Figure S2-F13	pKW18756-EWTL	Figure S4-N1	V47E/L199T
91	Figure S2-F14	pKW18756-FWEL	Figure S4-N2	V47F/L199E
92	Figure S2-F15	pKW18756-FWFL	Figure S4-N3	V47F/L199F
93	Figure S2-F16	pKW18756-FWHL	Figure S4-N4	V47F/L199H
94	Figure S2-F17	pKW18756-FWPL	Figure S4-N5	V47F/L199P
95	Figure S2-F18	pKW18756-FWQL	Figure S4-N6	V47F/L199Q
96	Figure S2-F19	pKW18756-GICS	Unstable	V47G/W185I/L199C/L210S
97	Figure S2-F20	pKW18756-GNNG	Unstable	V47G/W185N/L199N/L210G
98	Figure S2-F21	pKW18756-GWMD	Unstable	V47G/L199M/L210D
99	Figure S2-F22	pKW18756-HTSN	Unstable	V47H/W185T/L199S/L210N
100	Figure S2-F23	pKW18756-HWAL	Figure S4-N7	V47H/L199A

101	Figure S2-F24	pKW18756-HWCL	Figure S4-O1	V47H/L199C
102	Figure S2-F25	pKW18756-HWFL	Figure S4-O2	V47H/L199F
103	Figure S2-F26	pKW18756-HWGL	Figure S4-O3	V47H/L199G
104	Figure S2-F27	pKW18756-HWHL	Figure S4-O4	V47H/L199H
105	Figure S2-F28	pKW18756-HWKL	Figure S4-O5	V47H/L199K
106	Figure S2-F29	pKW18756-HWPL	Figure S4-O6	V47H/L199P
107	Figure S2-F30	pKW18756-HWQL	Figure S4-O7	V47H/L199Q
108	Figure S2-F31	pKW18756-HWTL	Figure S4-P1	V47H/L199T
109	Figure S2-F32	pKW18756-INKM	Unstable	V47I/W185N/L199K/L210M
110	Figure S2-F33	pKW18756-ISPP	Unstable	V47I/W185S/L199P/L210P
111	Figure S2-F34	pKW18756-KKGF	Unstable	V47K/W185K/L199G/L210F
112	Figure S2-F35	pKW18756-KQWT	Unstable	V47K/W185Q/L199W/L210T
113	Figure S2-F36	pKW18756-KWDL	Figure S4-P2	V47K/L199D
114	Figure S2-F37	pKW18756-KWML	Figure S4-P3	V47K/L199M
115	Figure S2-F38	pKW18756-LGCI	Unstable	V47L/W185G/L199C/L210I
116	Figure S2-F39	pKW18756-LQKQ	Unstable	V47L/W185Q/L199Q/L210K
117	Figure S2-F40	pKW18756-LWAM	Figure S4-P4	V47L/L199A/L210M
118	Figure S2-F41	pKW18756-LWGM	Figure S4-P5	V47L/L199G/L210M
119	Figure S2-F42	pKW18756-LWPM	Figure S4-P6	V47L/L199P/L210M
120	Figure S2-F43	pKW18756-LWQL	Figure S4-P7	V47L/L199Q
121	Figure S2-F44	pKW18756-LWSM	Figure S4-Q1	V47L/L199S/L210M
122	Figure S2-F45	pKW18756-MKIA	Unstable	V47M/W185K/L199I/L210A
123	Figure S2-F46	pKW18756-MLPE	Figure S4-Q2	V47M/W185L/L199P/L210E
124	Figure S2-F47	pKW18756-MWCL	Figure S4-Q3	V47M/L199C
125	Figure S2-F48	pKW18756-MWGA	Figure S4-Q4	V47M/L199G/L210A
126	Figure S2-F49	pKW18756-MWGC	Figure S4-Q5	V47M/L199G/L210C
127	Figure S2-F50	pKW18756-MWGL	Figure S4-Q6	V47M/L199G
128	Figure S2-F51	pKW18756-MWGM	Figure S4-Q7	V47M/L199G/L210M
129	Figure S2-F52	pKW18756-MWGS	Figure S4-Q8	V47M/L199G/L210S
130	Figure S2-F53	pKW18756-MWMM	Figure S4-R1	V47M/L199M/L210M
131	Figure S2-F54	pKW18756-MWNM	Figure S4-R2	V47M/L199N/L210M
132	Figure S2-F55	pKW18756-MWPA	Figure S4-R3	V47M/L199P/L210A
133	Figure S2-F56	pKW18756-MWPL	Figure S4-R4	V47M/L199P
134	Figure S2-F57	pKW18756-MWPM	Figure S4-R5	V47M/L199P/L210M
135	Figure S2-F58	pKW18756-MWSM	Figure S4-R6	V47M/L199S/L210M

136	Figure S2-F59	pKW18756-MWTL	Figure S4-S1	V47M/L199T
137	Figure S2-F60	pKW18756-MWTM	Figure S4-S2	V47M/L199T/L210M
138	Figure S2-F61	pKW18756-NWNL	Figure S4-S3	V47N/L199N
139	Figure S2-F62	pKW18756-PAVH	Unstable	V47P/W185A/L199V/L210H
140	Figure S2-F63	pKW18756-PDQE	Unstable	V47P/W185D/L199Q/L210E
141	Figure S2-F64	pKW18756-PSMF	Unstable	V47P/W185S/L199M/L210F
142	Figure S2-F65	pKW18756-QWPM	Figure S4-S4	V47Q/L199P/L210M
143	Figure S2-F66	pKW18756-RESW	Unstable	V47R/W185E/L199S/L210W
144	Figure S2-F67	pKW18756-RVCR	Unstable	V47R/W185V/L199C/L210R
145	Figure S2-F68	pKW18756-SALG	Unstable	V47S/W185A/L210G
146	Figure S2-F69	pKW18756-SDVK	Unstable	V47S/W185D/L199V/L210K
147	Figure S2-F70	pKW18756-SRST	Unstable	V47S/W185R/L199S/L210T
148	Figure S2-F71	pKW18756-SWAL	Figure S4-S5	V47S/L199A
149	Figure S2-F72	pKW18756-SWRL	Figure S4-S6	V47S/L199R
150	Figure S2-F73	pKW18756-SWYL	Figure S4-S7	V47S/L199Y
151	Figure S2-F74	pKW18756-TGKH	Unstable	V47T/W185G/L199K/L210H
152	Figure S2-F75	pKW18756-TWAL	Figure S4-S8	V47T/L199A
153	Figure S2-F76	pKW18756-TWHL	Figure S4-T1	V47T/L199H
154	Figure S2-F77	pKW18756-TWRL	Figure S4-T2	V47T/L199R
155	Figure S2-F78	pKW18756-TWWL	Figure S4-T3	V47T/L199W
156	Figure S2-F79	pKW18756-WWAL	Figure S4-T4	V47W/L199A
157	Figure S2-F80	pKW18756-WWKI	Unstable	V47W/L199K/L210I
158	Figure S2-F81	pKW18756-WWPA	Figure S4-T5	V47W/L199P/L210A
159	Figure S2-F82	pKW18756-WWPL	Figure S4-T6	V47W/L199P
160	Figure S2-F83	pKW18756-WWPW	Figure S4-U1	V47W/L199P/L210W
161	Figure S2-F84	pKW18756-WWQI	Unstable	V47W/L199Q/L210I
162	Figure S2-F85	pKW18756-WWQL	Figure S4-U2	V47W/L199Q
163	Figure S2-F86	pKW18756-WWRL	Figure S4-U3	V47W/L199R
164	Figure S2-F87	pKW18756-YQCG	Unstable	V47Y/W185Q/L199C/L210G
165	Figure S2-F88	pKW18756-YYMD	Unstable	V47Y/W185Y/L199M/L210D



**Figure 2-7-3.** 初期の学習用ライブラリー作製のベクターマップ

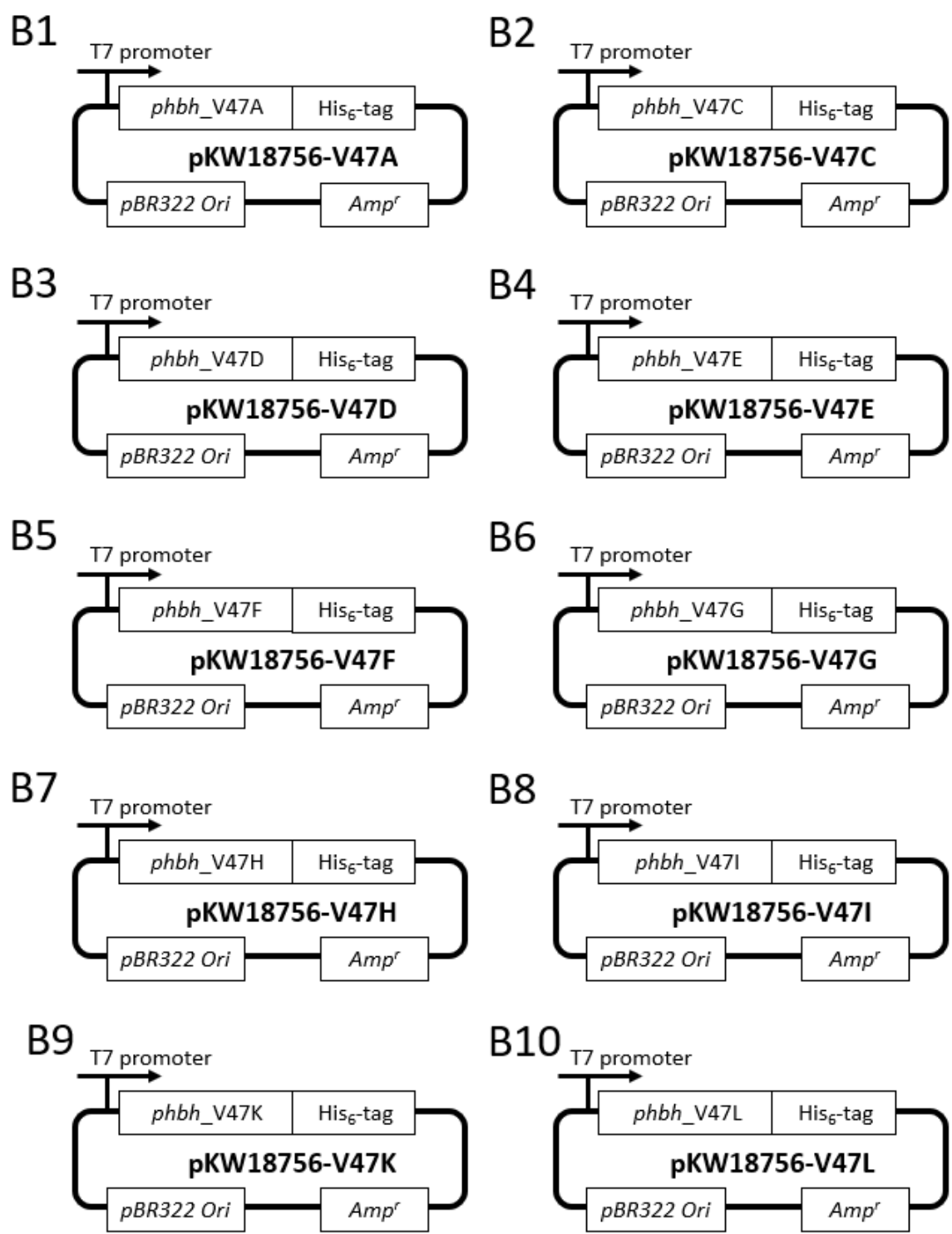


Figure 2-7-3. (前ページの続き)

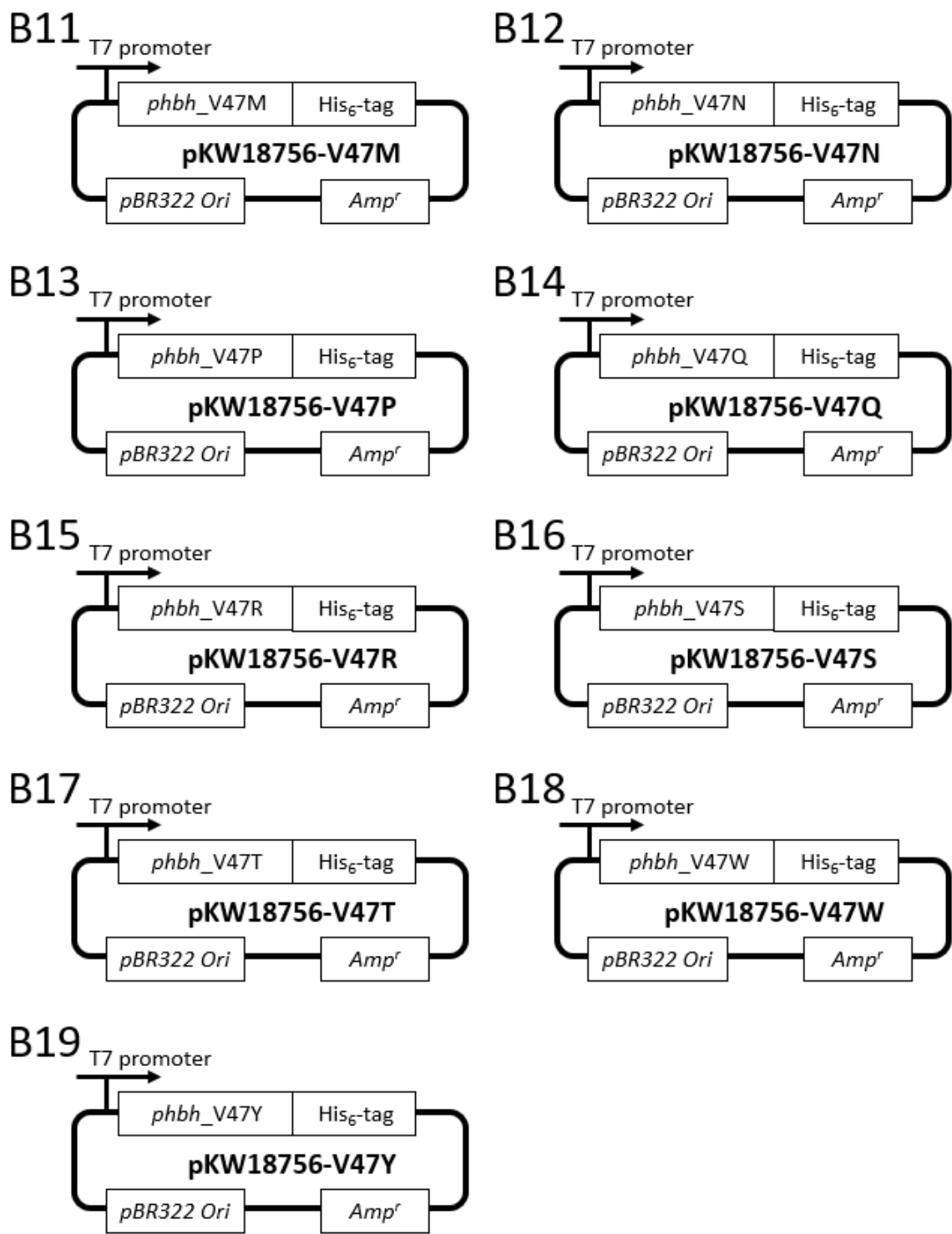


Figure 2-7-3. (前ページの続き)

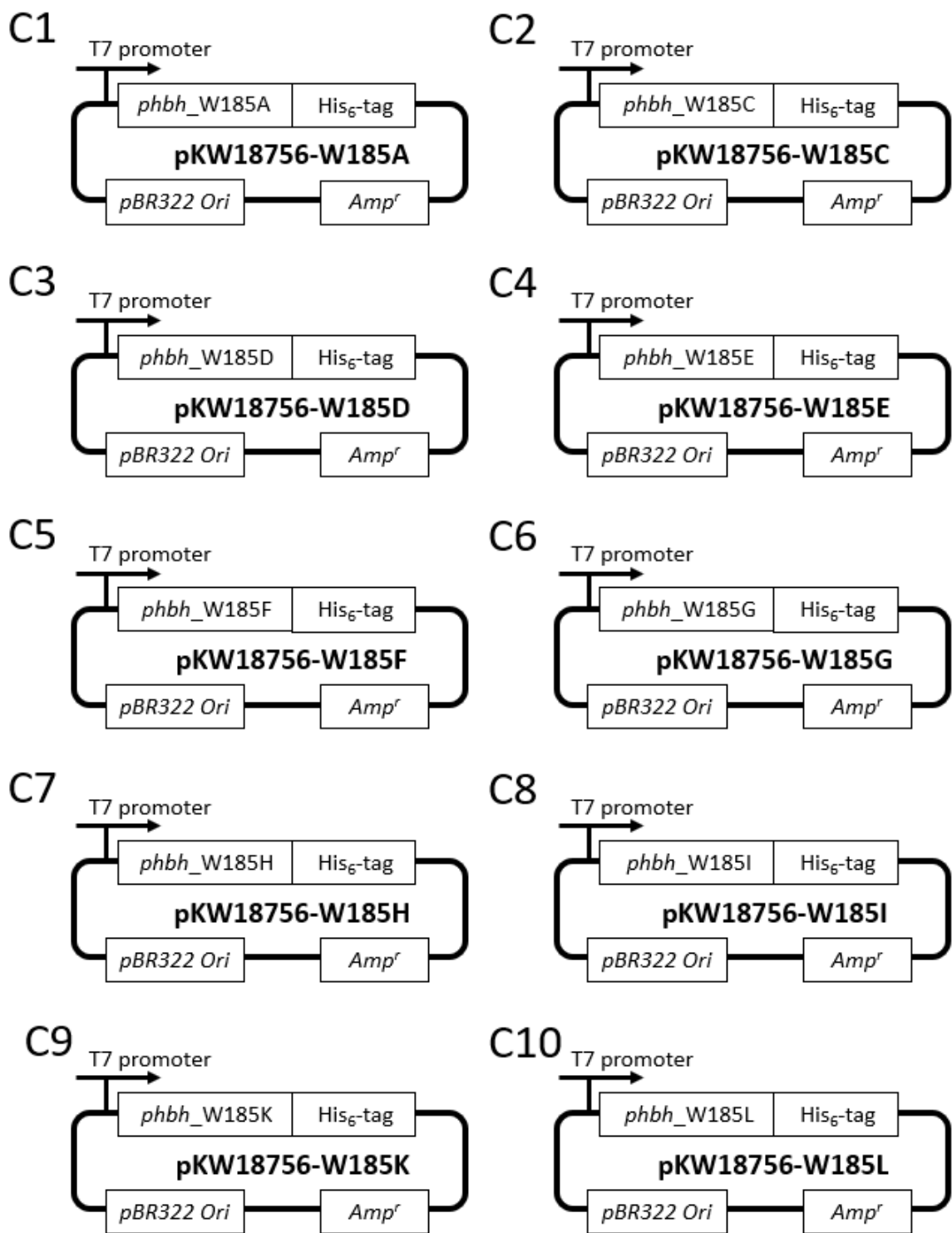
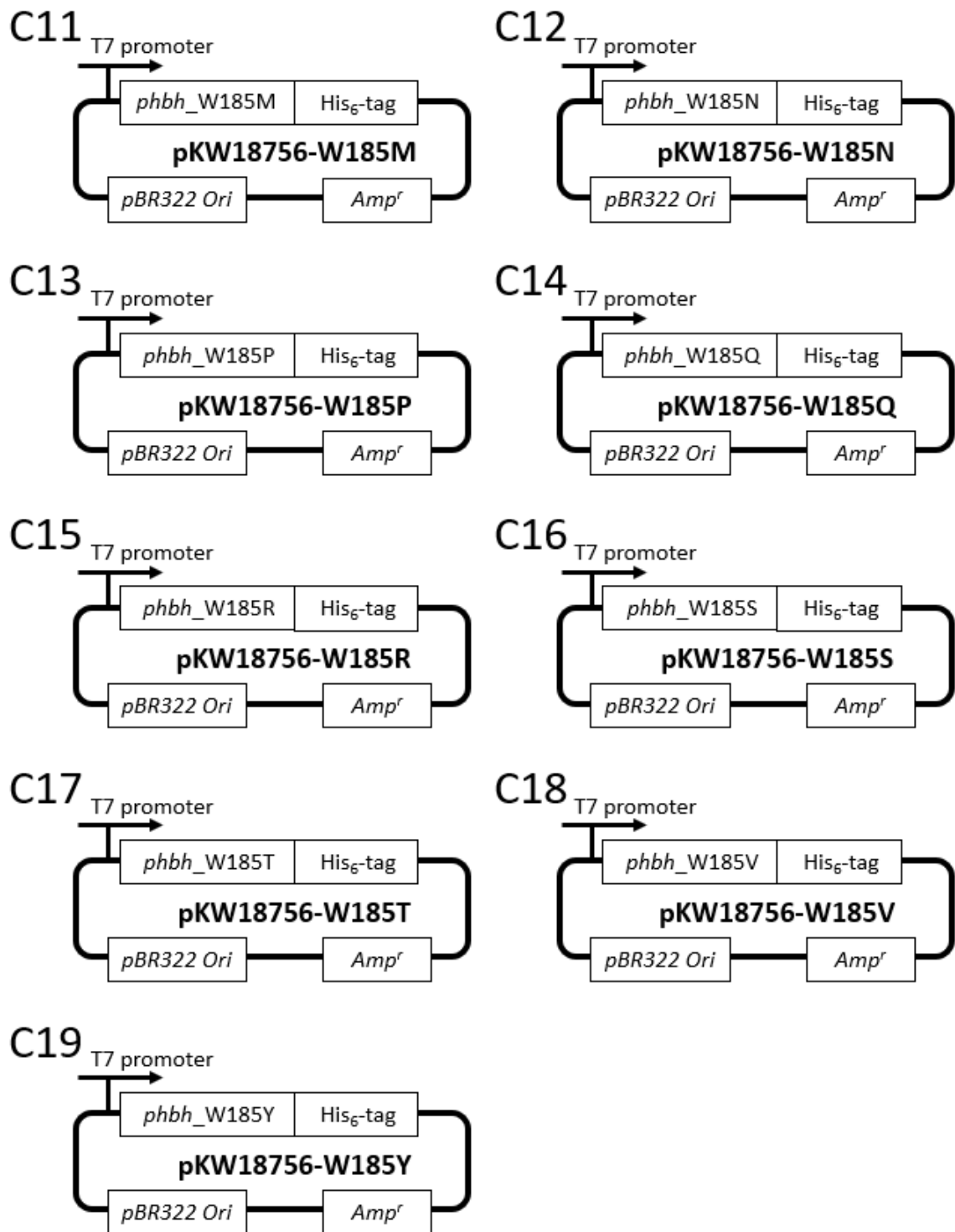


Figure 2-7-3. (前ページの続き)



**Figure 2-7-3.** (前ページの続き)



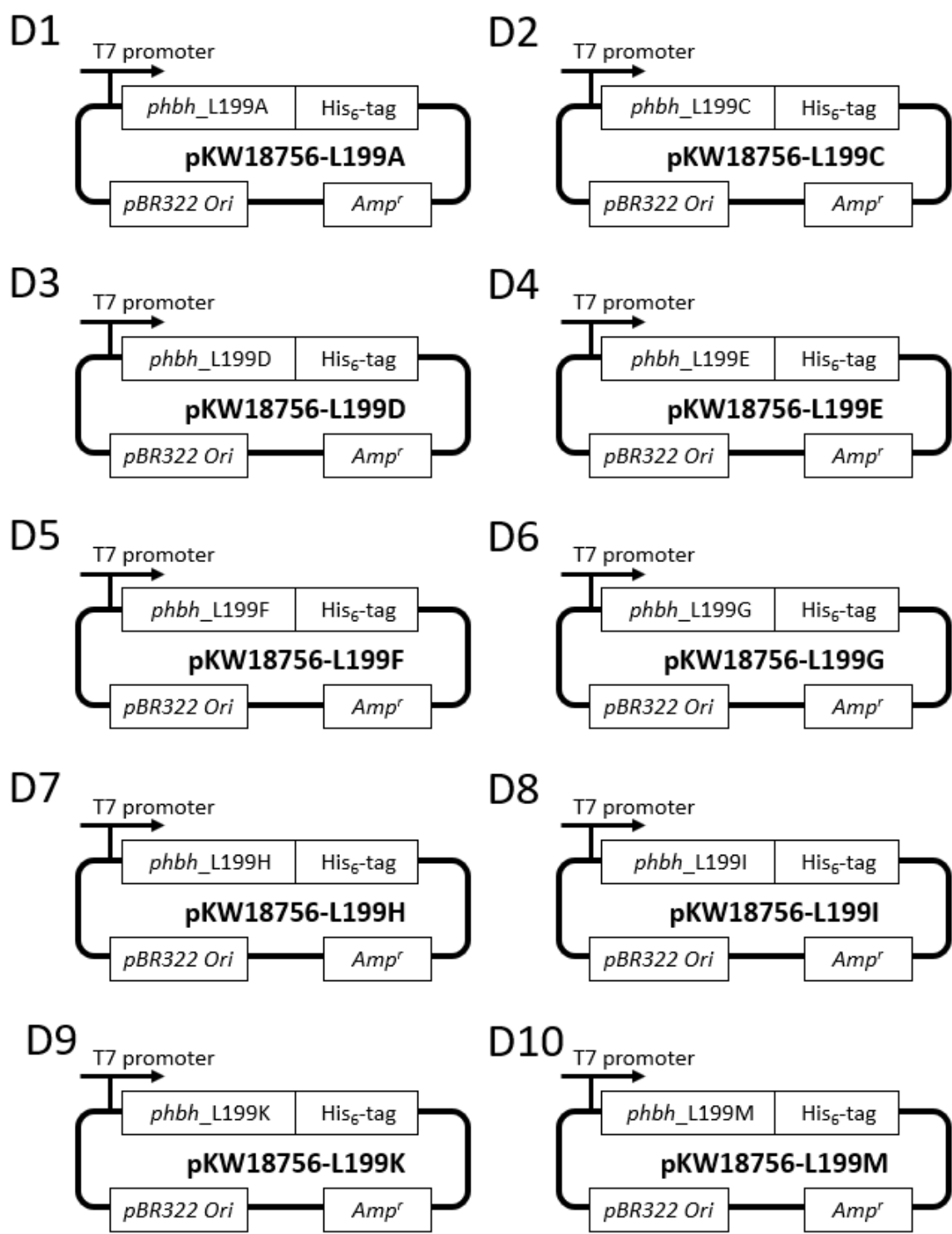


Figure 2-7-3. (前ページの続き)

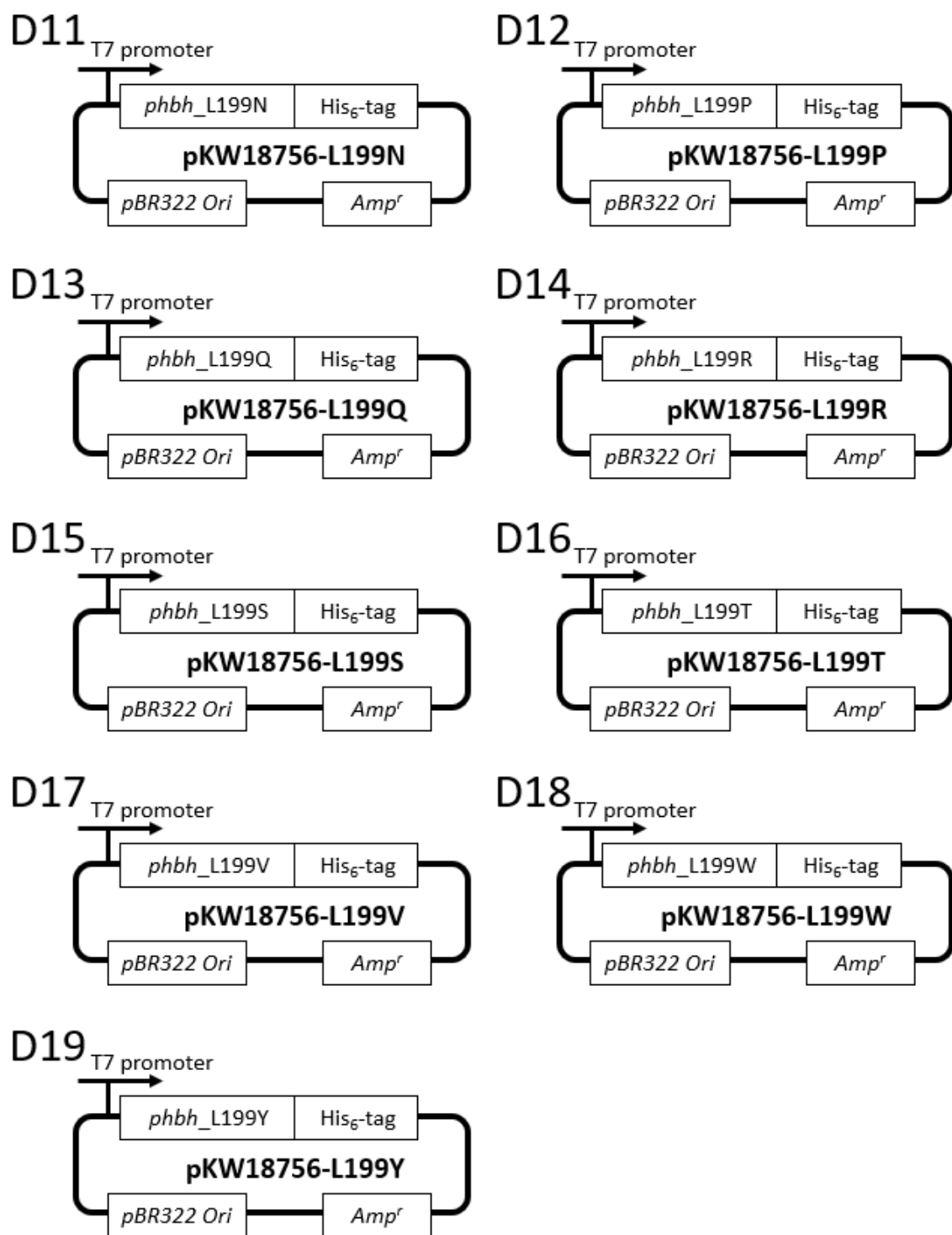


Figure 2-7-3. (前ページの続き)

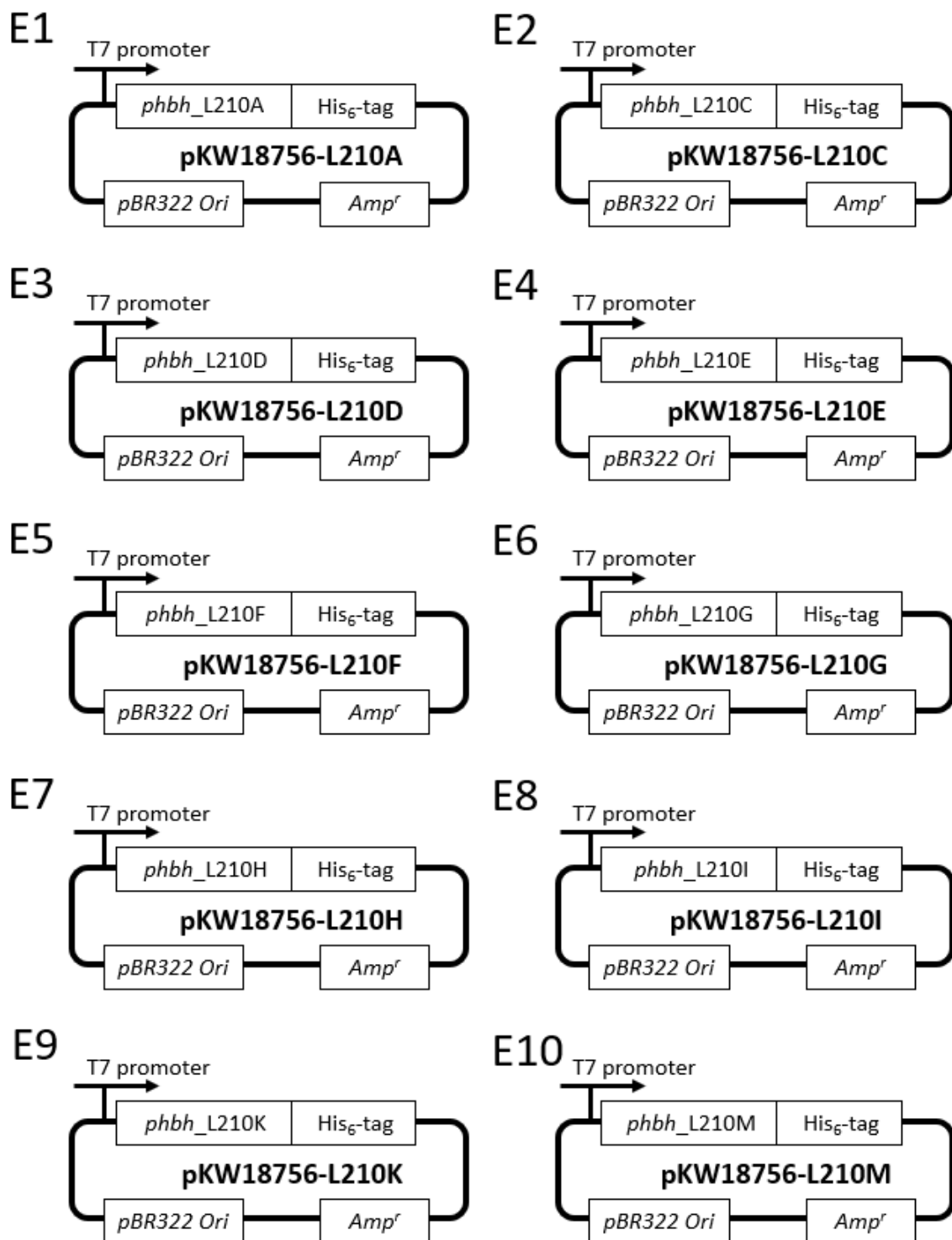


Figure 2-7-3. (前ページの続き)

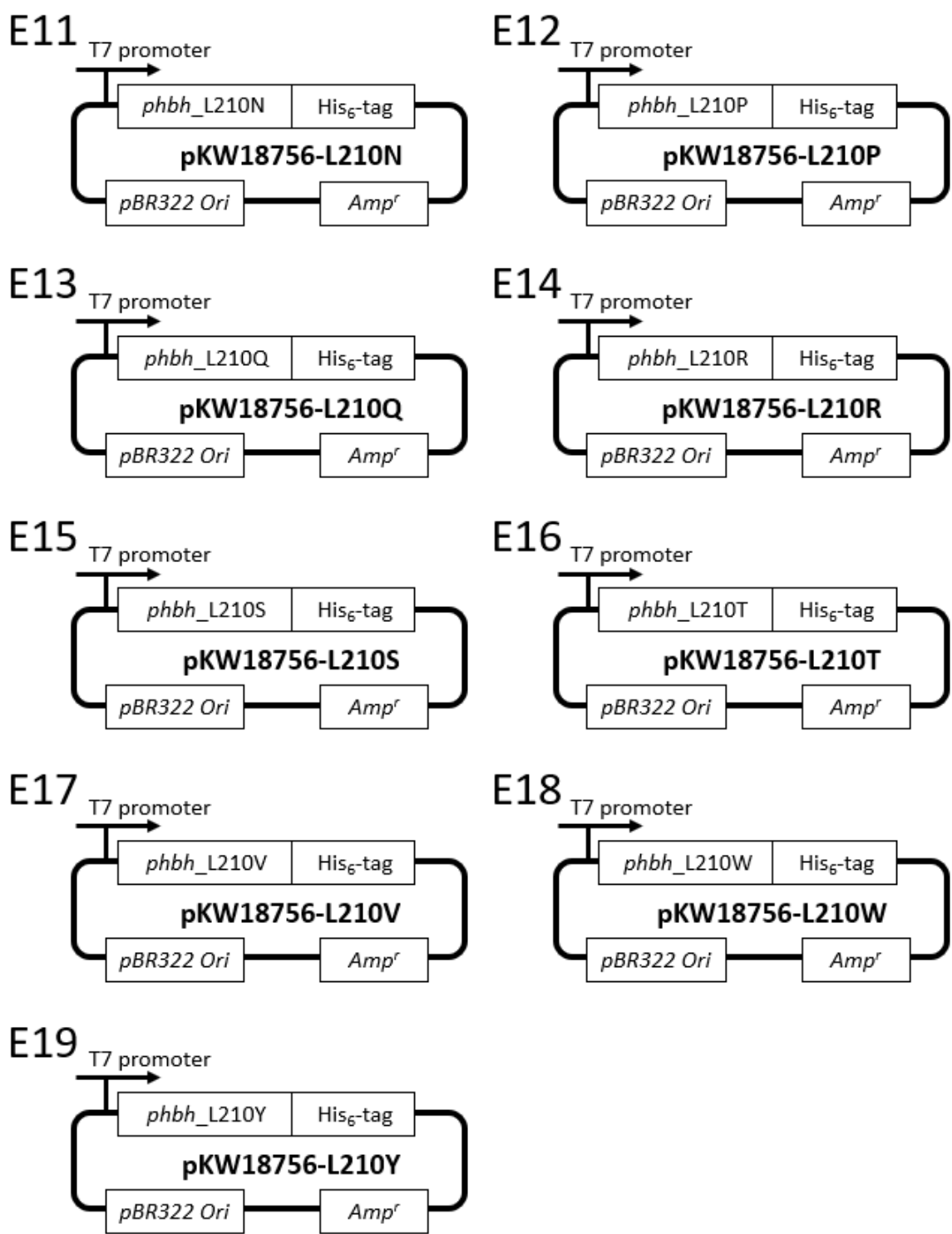


Figure 2-7-3. (前ページの続き)

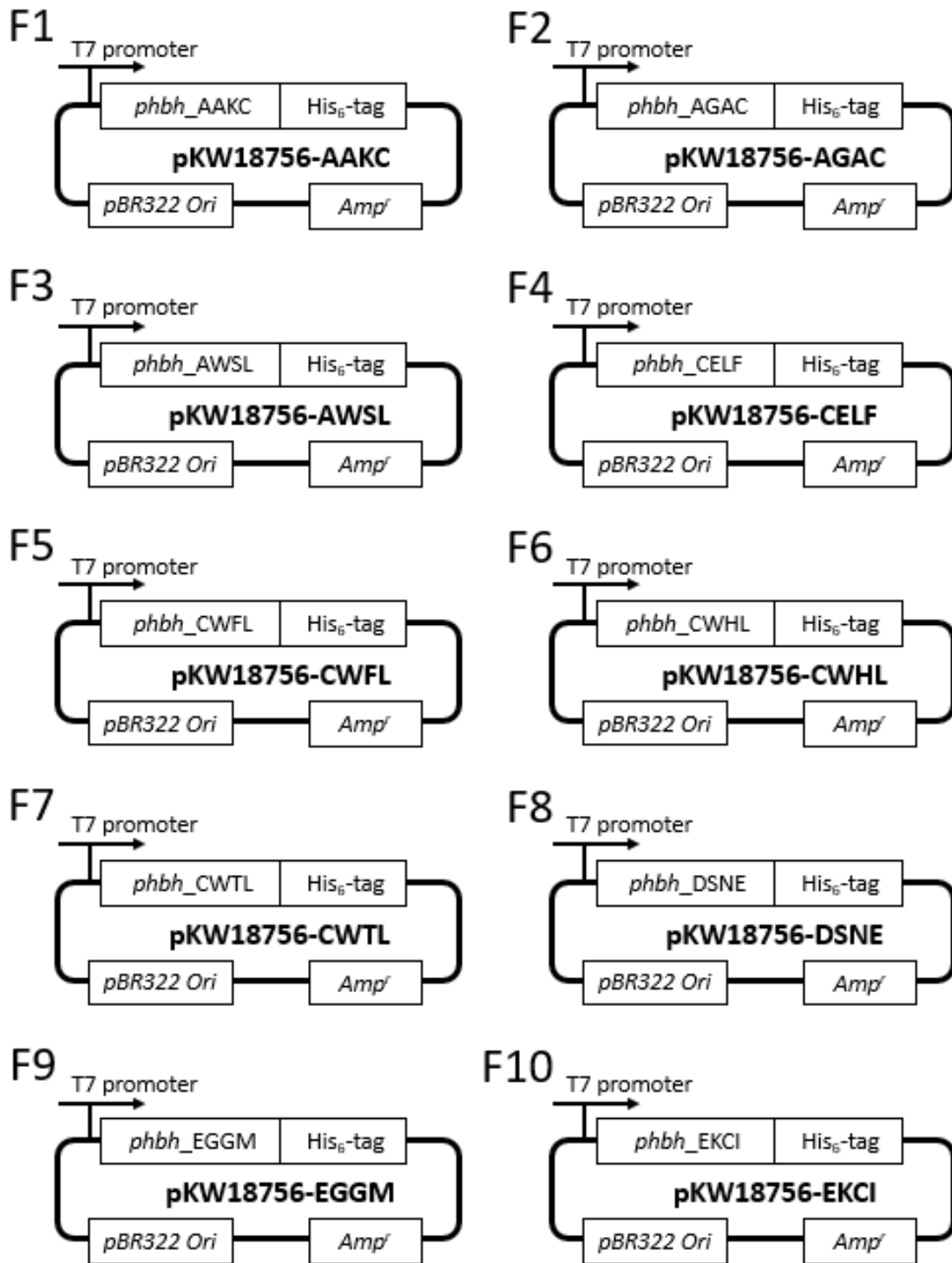


Figure 2-7-3. (前ページの続き)

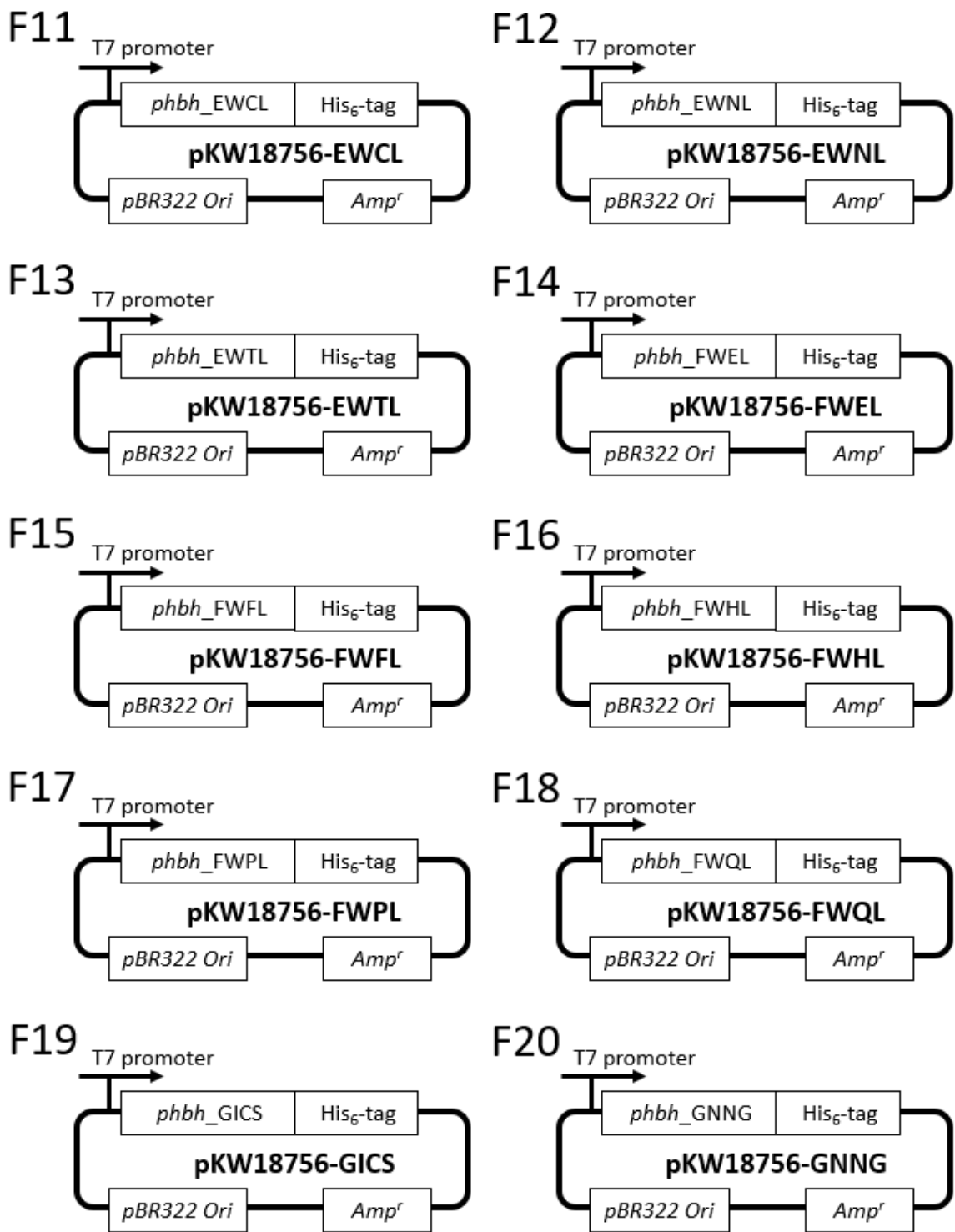


Figure 2-7-3. (前ページの続き)

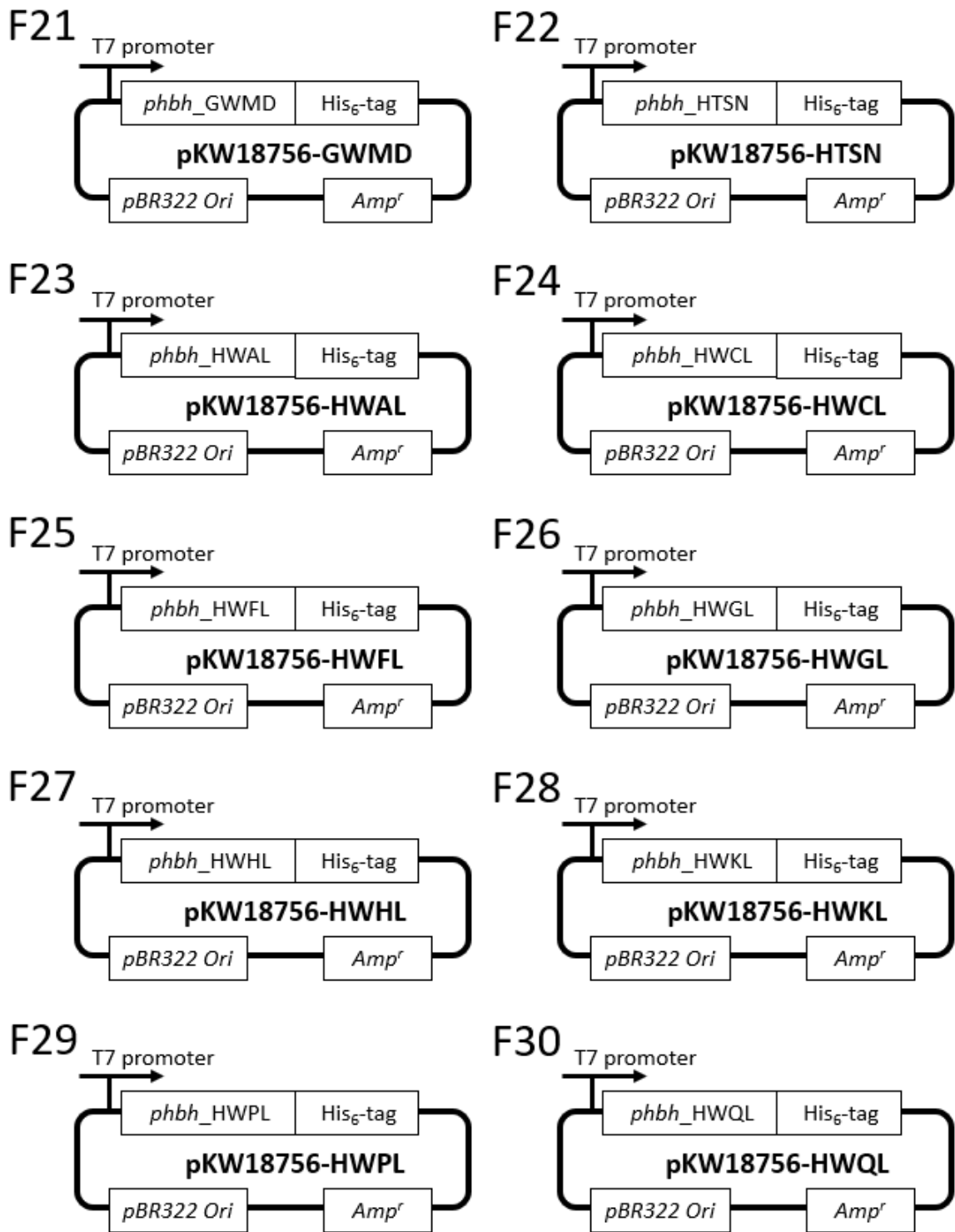


Figure 2-7-3. (前ページの続き)

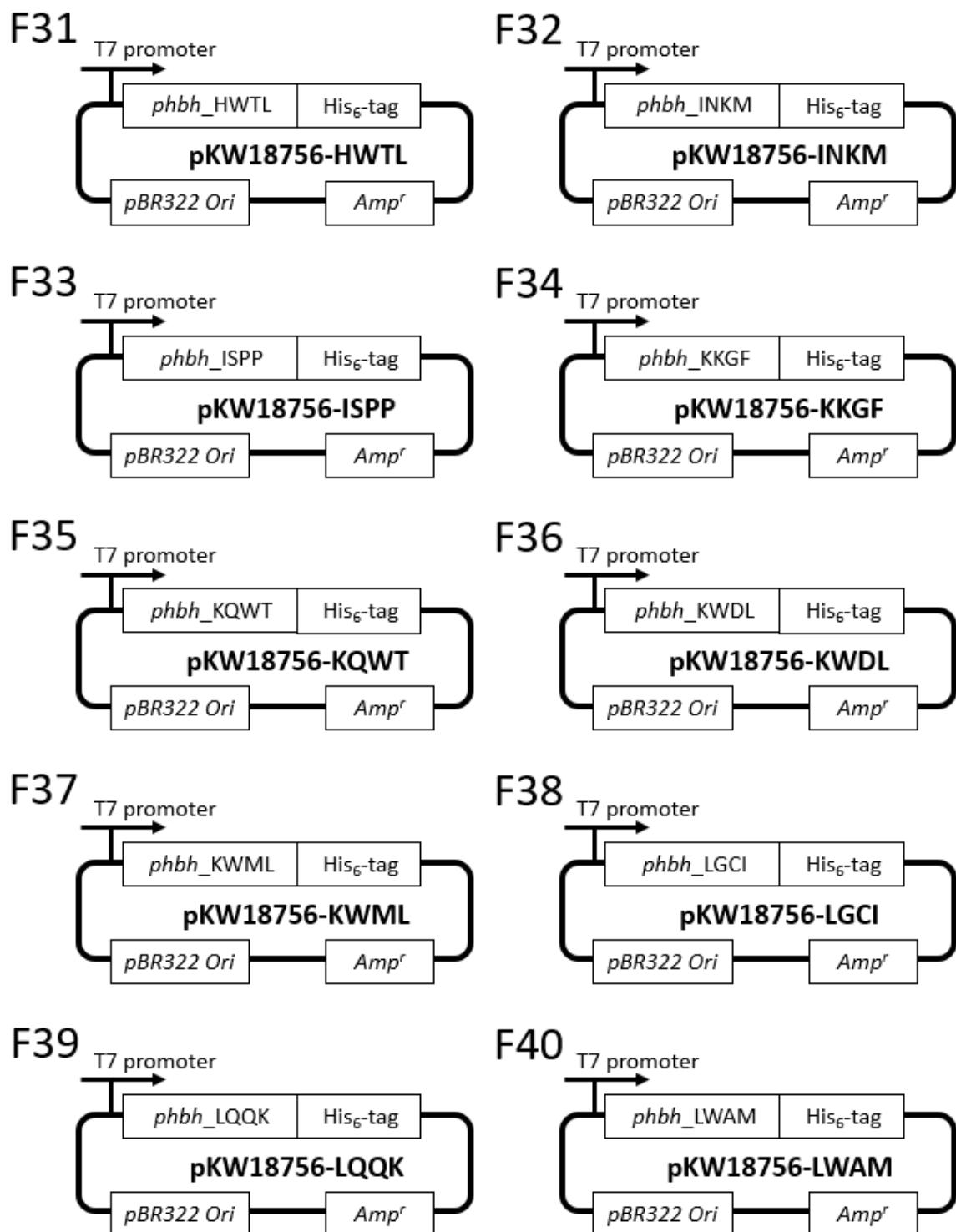


Figure 2-7-3. (前ページの続き)



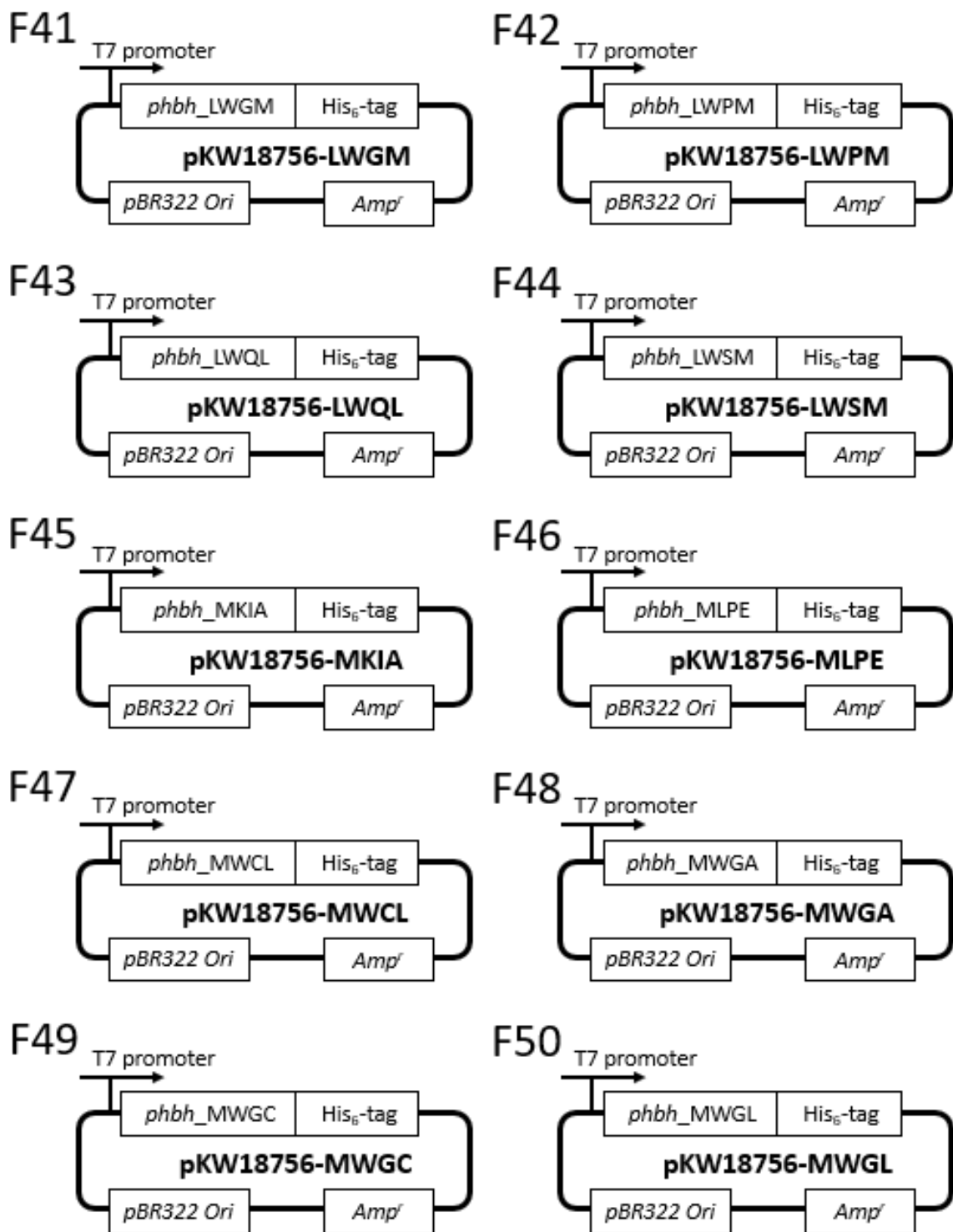


Figure 2-7-3. (前ページの続き)

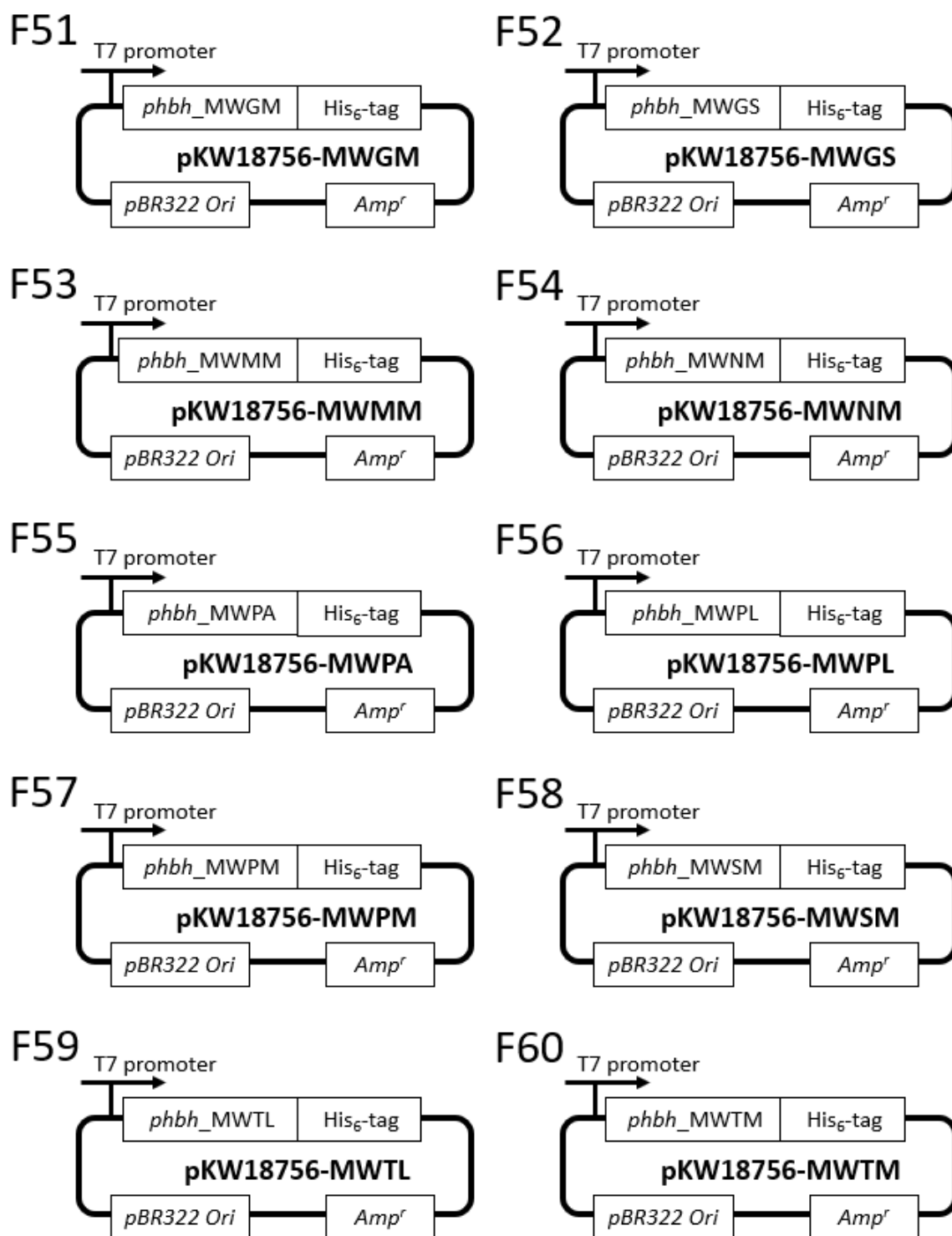


Figure 2-7-3. (前ページの続き)

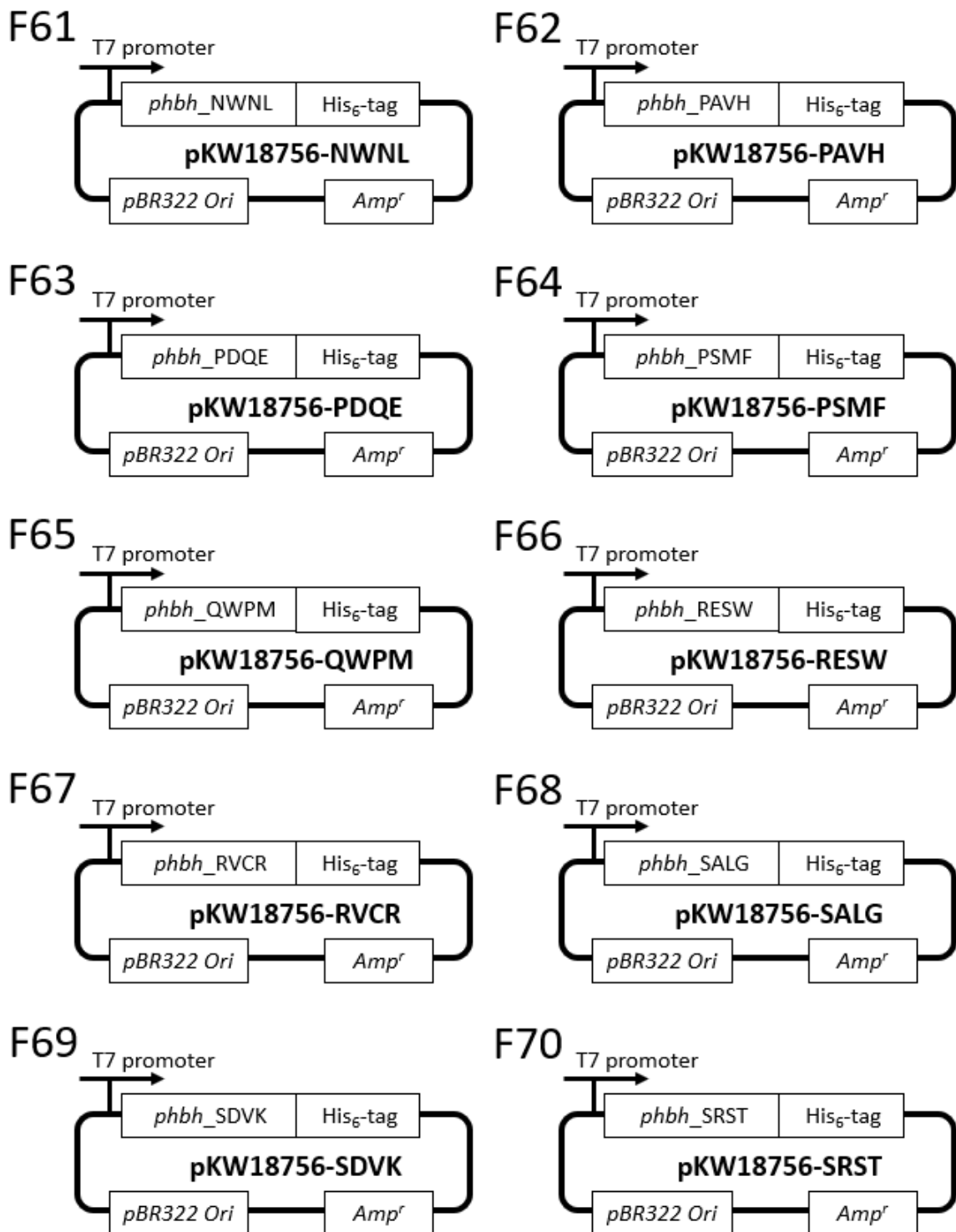


Figure 2-7-3. (前ページの続き)

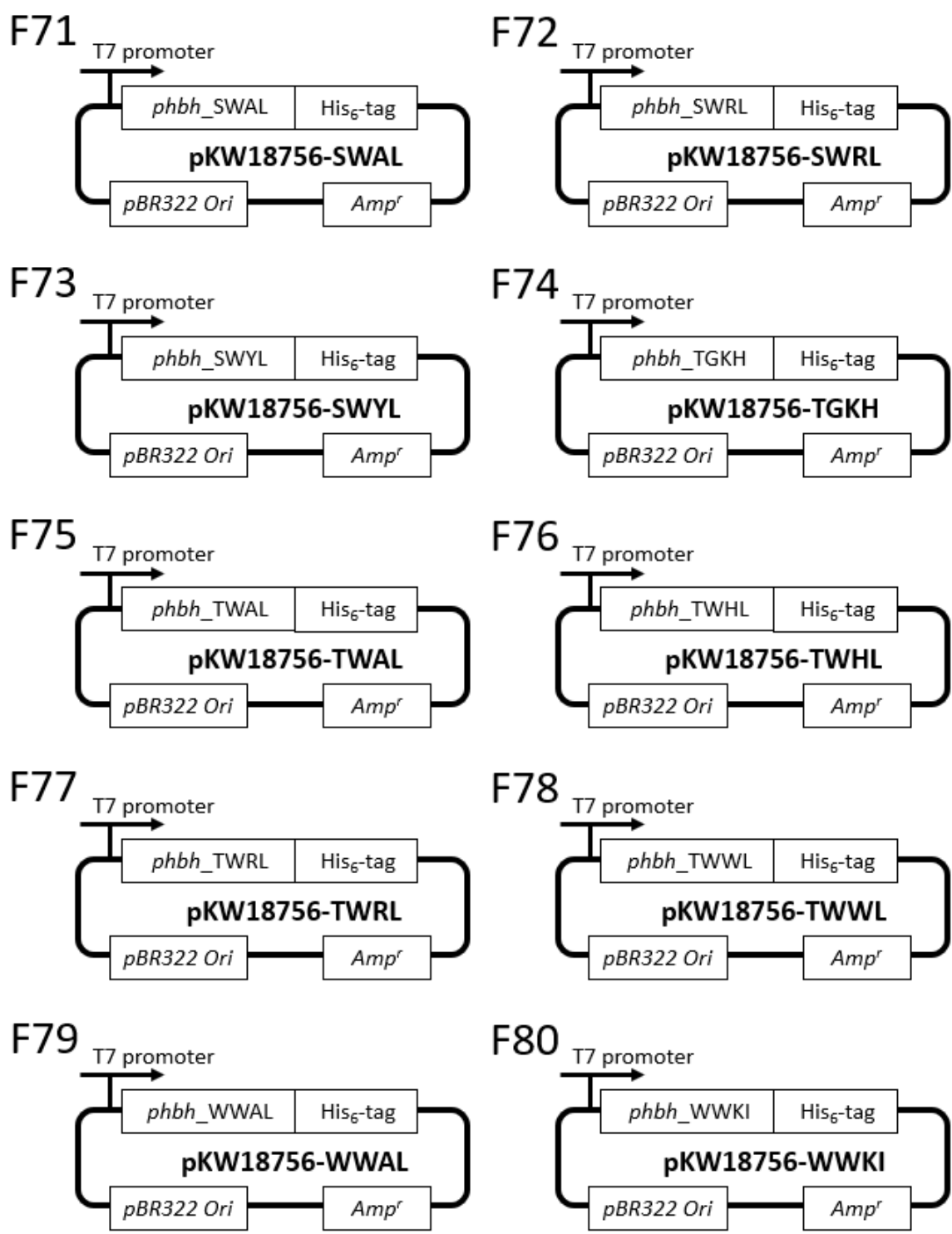


Figure 2-7-3. (前ページの続き)

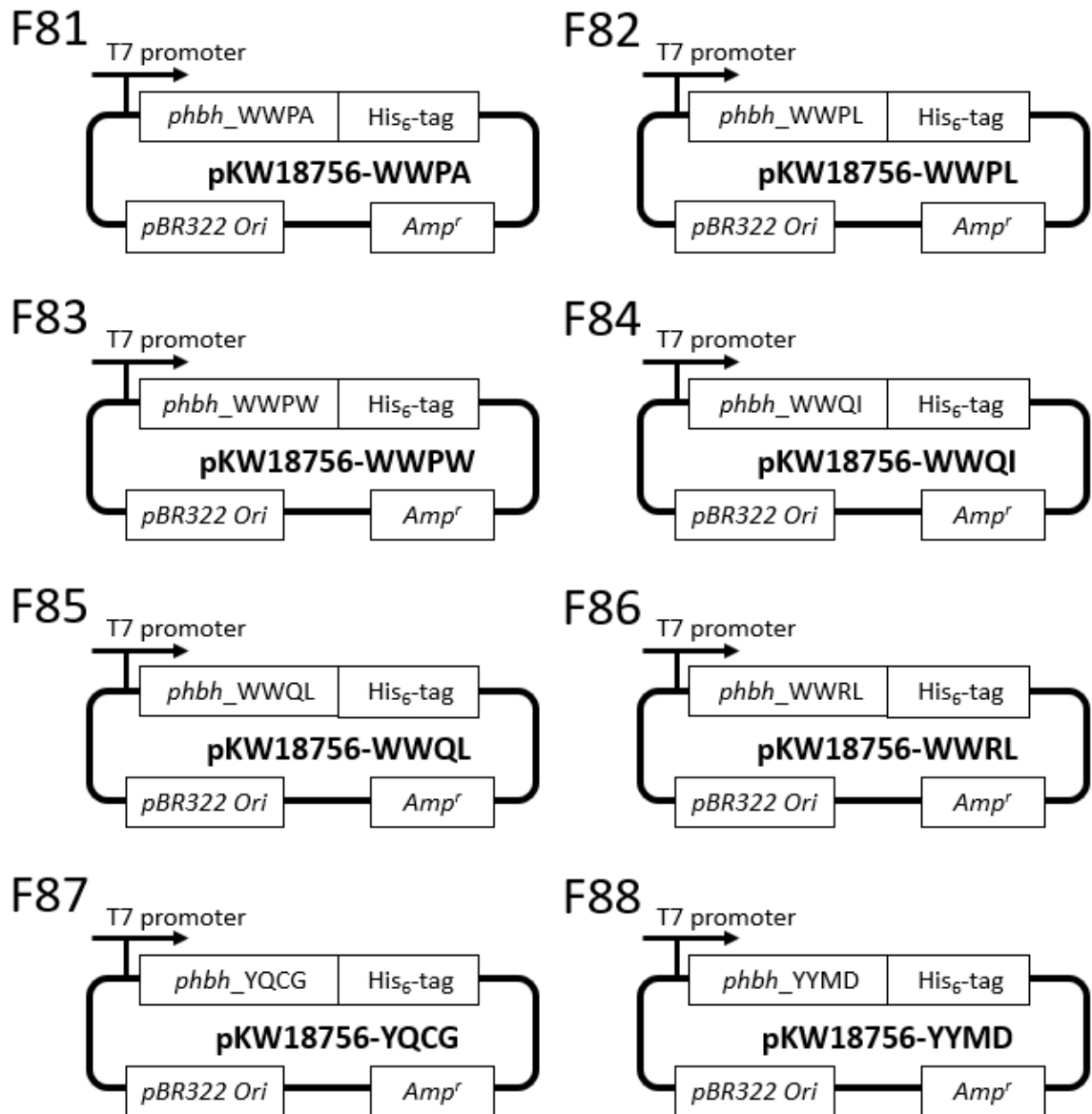
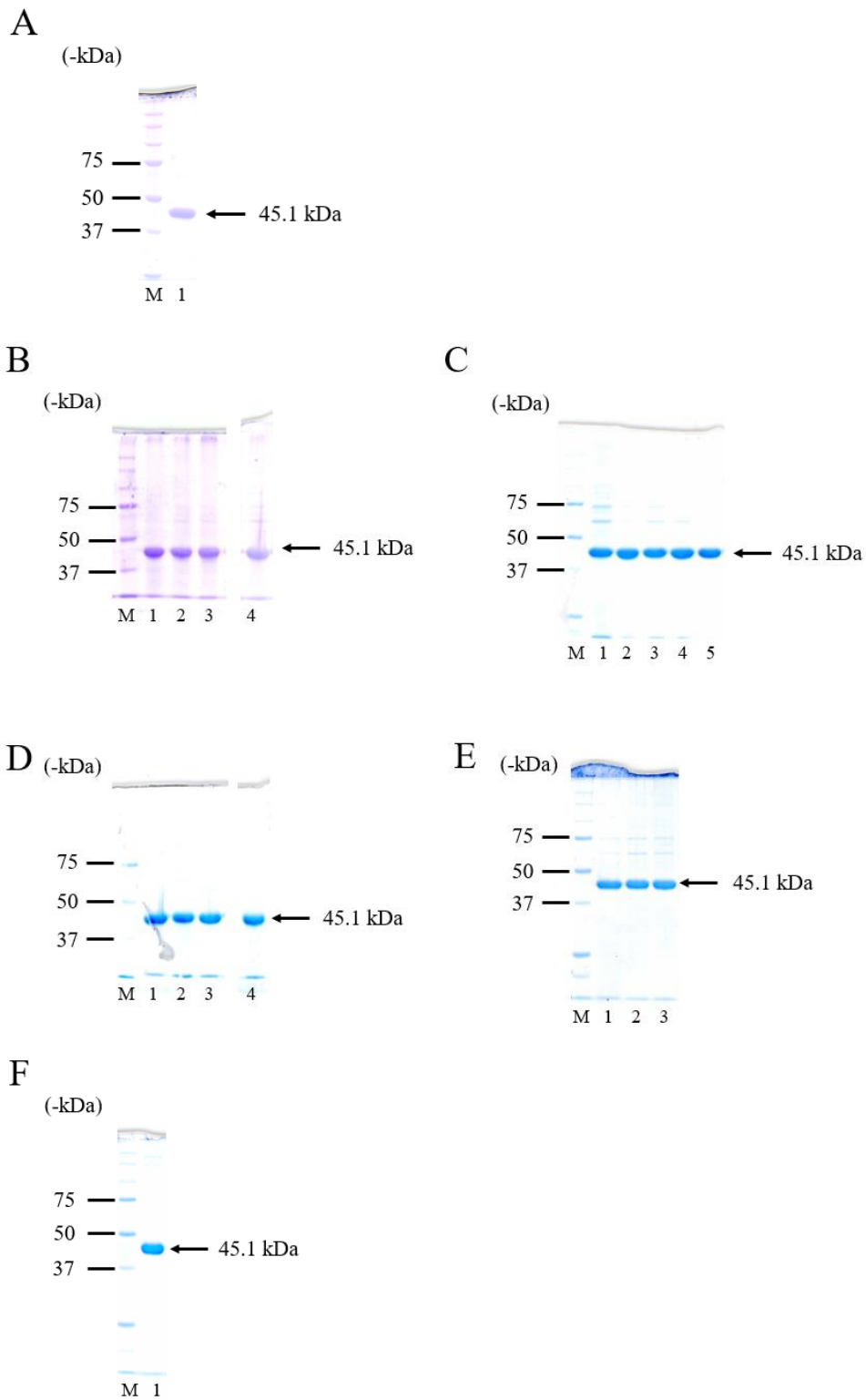
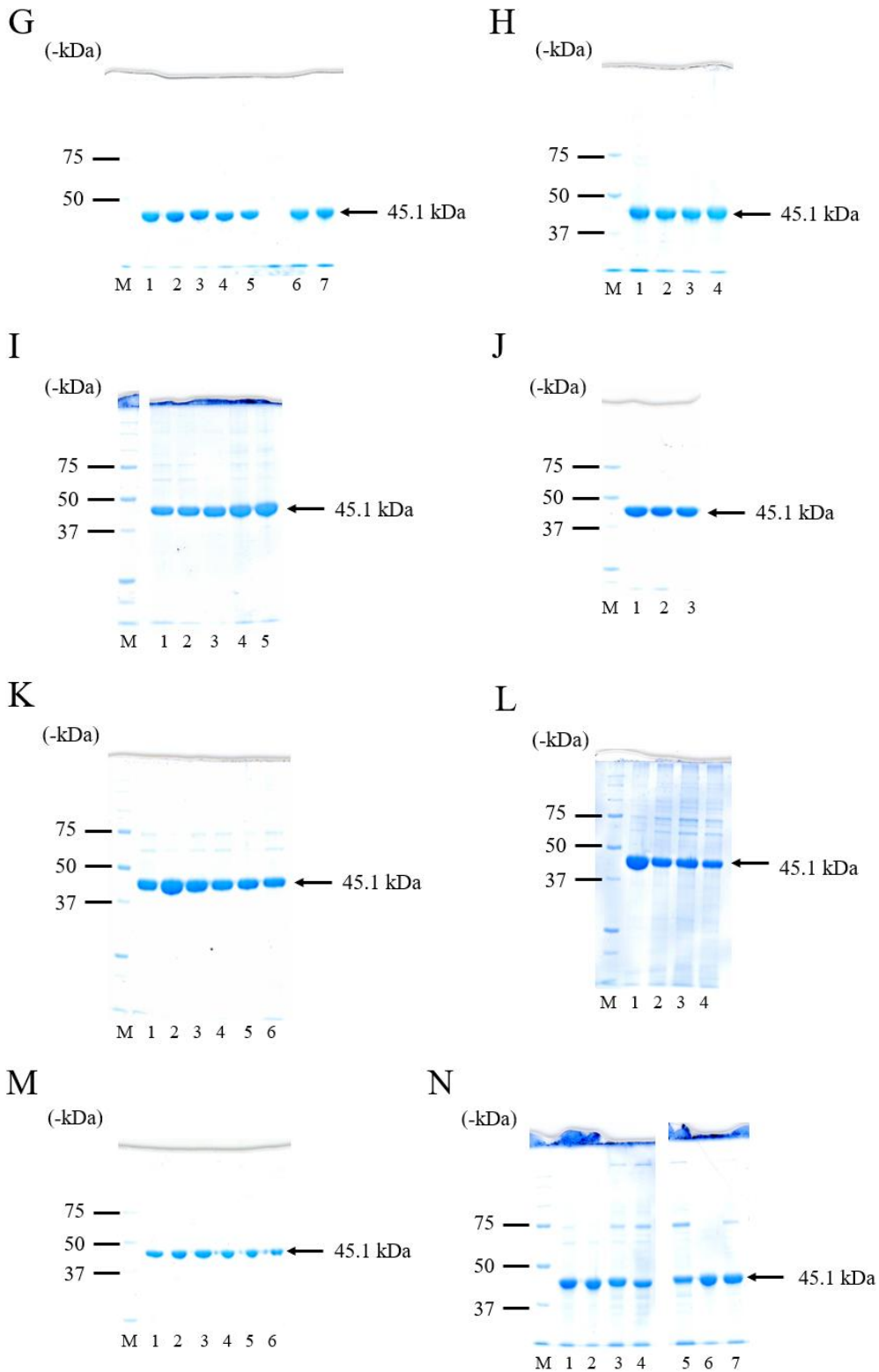


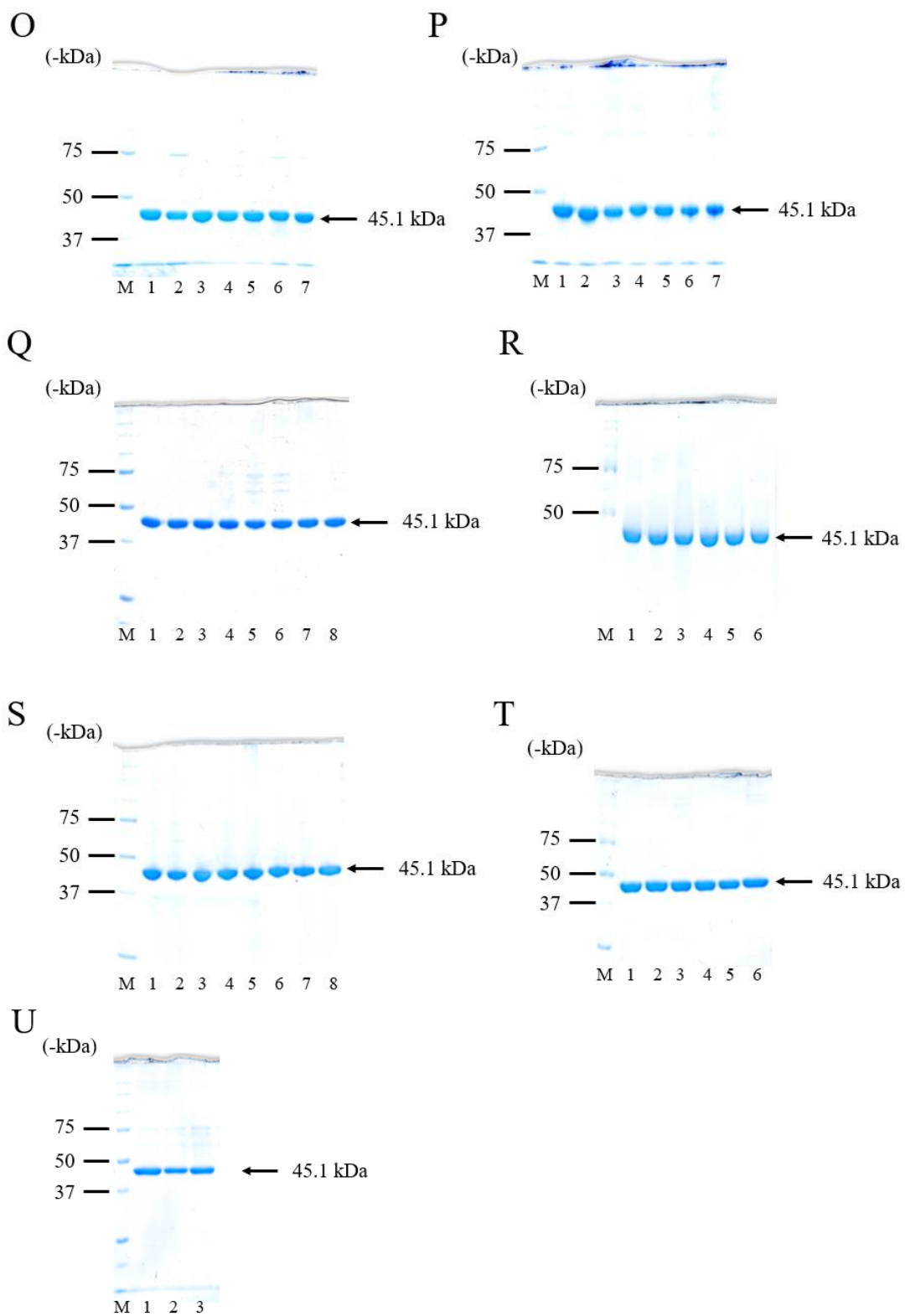
Figure 2-7-3. (前ページの続き)



**Figure 2-7-4.** 初期ライブラリーの変異体の SDS-PAGE



**Figure 2-7-4.** (前ページの続き)



**Figure 2-7-4.** (前ページの続き)



## 第八節 引用文献

- (1) Gatti, D. L.; Entsch, B.; Ballou, D. P.; Ludwig, M. L. pH-Dependent Structural Changes in the Active Site of *p*-Hydroxybenzoate Hydroxylase Point to the Importance of Proton and Water Movements during Catalysis. *Biochemistry* **1996**, *35*, 567-578.
- (2) Gran-Scheuch, A.; Parra, L.; Fraaije, M. W. Systematic Assessment of Uncoupling in Flavoprotein Oxidases and Monooxygenases. *ACS Sustainable Chem. Eng.* **2023**, *11*, 4948–4959.
- (3) Benítez, A. R.; Tweedy, S.; Baker-Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L. 3rd.; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H. Structural basis for selectivity in flavin-dependent monooxygenase-catalyzed oxidative dearomatization. *ACS Catal.* **2019**, *9*, 3633–3640.

## 第三章 機械学習による高活性変異体の探索

### 第一節 機械学習の概要

第二章で収集した活性試験のデータを用いて、機械学習に着手した。今回は、活性部位を構成する4種類の疎水性アミノ酸残基（Val47, Trp185, Leu199, Leu210）の変異の種類と活性の強さを教師データとして機械学習をすることにした。アミノ酸残基の種類は文字列であり、それに対して酵素の活性の強さは数字である。両者は異なる種類の情報であるため、それらを統合するためにはアミノ酸配列を数字に変換する必要がある。また、機械学習の相関関数を作成する際にも、使用するモデルによって機械学習の結果が変わってくることが予想される。よって、教師データを用いて小規模な機械学習をすることで、適切なモデル関数の構築を目指した。

## 第二節 Protein descriptor による特徴ベクトルの作成

酵素を構成するアミノ酸残基は 20 種類存在し、それぞれ異なった立体障害、極性、疎水性、酸性度、塩基性度、分子量などの物性値を持つ。これらの物性値は酵素の基質認識性において非常に重要であるが、少ない種類の物性値をもとに特徴量を計算すると機械学習の結果が偏った結果になってしまう。一方、膨大な物性値の情報をそのまま機械学習に用いてしまうと教師データの持つ情報量が膨大になり、必要な計算リソースや計算時間も膨大なものになってしまう。これらの問題を解決して効率的な機械学習を行うためには、多くの種類の物性値をもとに集約したパラメーターを事前に計算する必要がある。そのために使用するのが Protein descriptor である。

Protein descriptor には多くの種類が知られており、代表的な物として ST-scale,<sup>1)</sup> Z-scale,<sup>2)</sup> T-scale,<sup>3)</sup> FASGAI,<sup>4)</sup> MS-WHIM,<sup>5)</sup> ProtFP,<sup>6)</sup> VHSE,<sup>7)</sup> BLOSUM<sup>8)</sup>などが存在する。Protein descriptor はそれぞれ異なった方法で計算されているため、使用する protein descriptor によって機械学習の結果も変わってくる。よって、適切な protein descriptor を使用することが、機械学習をする上でも重要になってくる。

上記の protein descriptor に加えて、タンパク質中の特定のアミノ酸残基について、他の類似酵素と比較をすることで、それらの進化的保存度を指標に計算された PSSM<sup>9)</sup>も知られてる。この PSSM は一次元のパラメーターであるので単独で使用することは難しいが、上記の protein descriptor と PSSM を組み合わせることで、protein descriptor 単体で機械学習をするよりも予測精度が上昇することがある。

### 第三節 自然言語処理用深層学習プログラム：BERT による特徴ベ

#### クトルの抽出

近年、深層学習を用いた酵素の機能改変が注目されており、その中で最も代表的な深層学習プログラムである BERT について説明する。BERT は Google 社とトロント大学のグループによって、自然言語処理のために開発された深層学習プログラムである。特徴ベクトルの抽出過程では transformer が用いられており、文字列の全長の中の一箇所だけを隠し、アテンションを用いて各文字の持つ「重み」を調整することで、残りの全ての文字列から隠された部分を予想できるように調節をする手法である masked language modeling (Figure 3-3-1) が用いられている。いわば「穴埋め問題」を解くような手法により、特徴量の抽出を行っている。

#### Masked language modeling

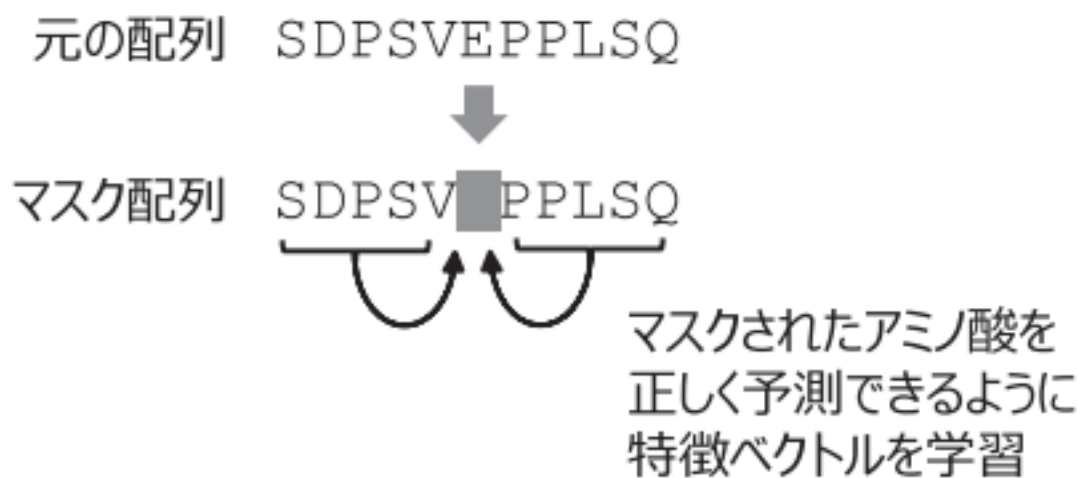
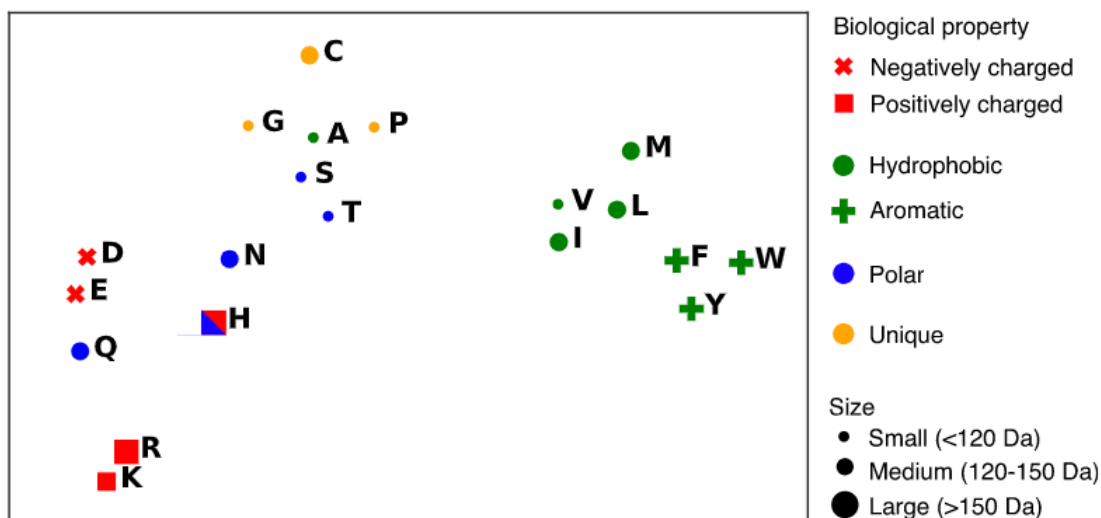


Figure 3-3-1. Transformer による特徴ベクトルの抽出過程。

しかし、transformer による特徴ベクトルの抽出には膨大な計算リソースが必要であり、より少ない計算リソースでの特徴量ベクトルの抽出を行うべく、小規模化されたモデルである TAPE transformer<sup>10)</sup>が開発された。さらには、Facebook の人工知能グループによる研究では、この TAPE transformer が酵素の機能改変においても有用であることが示された。その研究では、Pfam データベースに登録されている 250 万種類のタンパク配列に対して、上記のように TAPE transformer による特徴ベクトルの抽出をした。抽出された特徴ベクトル

に対して **t-SNE** を用いた次元削減をすると、似たような物性値を持つアミノ酸残基が近傍にグループ化され、**TAPE transformer** により抽出された特徴ベクトルがアミノ酸残基の物性値を反映していることが確認された (**Figure 3-3-2**)<sup>11)</sup>。

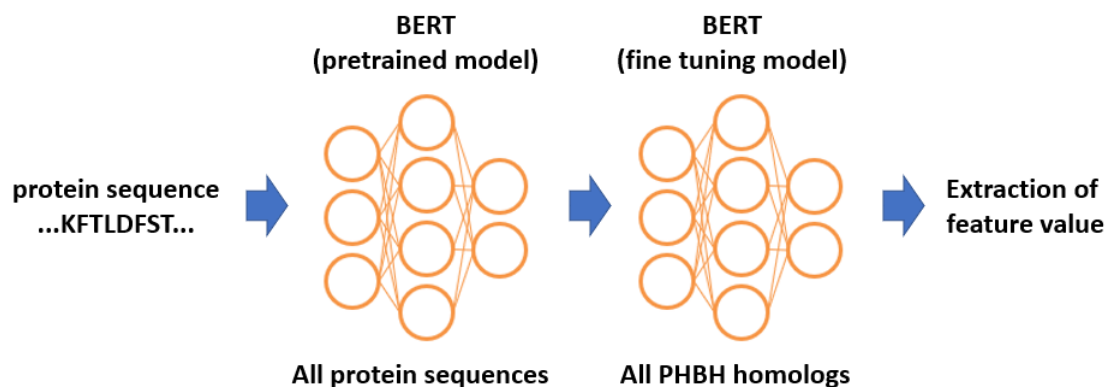


**Figure 3-3-2.** TAPE transformer により抽出された特徴ベクトルの t-SNE による次元削減

さらには、従来の **protein descriptor** の抽出過程では、人間が選択した各アミノ酸残基の物性値をもとにして計算されているため、物性値を選択する段階で人間による先入観が生じ、機械学習の予測結果にもバイアスが生じてしまう可能性がある。それに対して、**TAPE transformer** による特徴ベクトルの抽出では、データベースに登録されている膨大なタンパク質の配列を全て用いて学習することが可能になる。つまりは、人間により特徴量を選択する過程をスキップすることができるため、先入観なく特徴ベクトルを抽出することができる。

また、機械学習の領域では、まずは広範なデータを用いて大規模な学習（表現学習）を行い、その後自身が行う予定であるタスクに関連した情報を用いて、再度学習（転移学習）させることが一般的である。その理論を **BERT** にも応用したのが、産総研の齋藤らによって開発された **EvoTuned-BERT**<sup>12)</sup>である。その特徴ベクトルの抽出過程では、従来の **BERT** と同様にデータベースに登録されている全てのタンパク質配列をもとにして表現学習し、次に機能改変をしたいタンパク質配列と相溶性が高いタンパク質に絞って転移学習することで、より目的のタンパク質に適した特徴ベクトルが抽出される (**Figure 3-3-3**)。本研究では、**TAPE transformer** を用いて Pfam データベースに登録されている全タンパク質の配列を用いて、上述した **masked language modeling** による表現学習を

行った。そして、データベース上の  $Evalue < 0.001$  (データベース上のタンパク質の配列が全てランダムであると仮定したときに、データベース上に存在する PHBH と同じ配列の本数の期待値が  $0.001$  よりも小さい) であるタンパク質配列に対して同様に転移学習をすることで、より PHBH に適した特徴量ベクトルを抽出した。



**Figure 3-3-3.** Evtuned-BERT による特徴ベクトルの抽出過程。

## 第四節 Nested-クロスバリデーションによる精度評価

機械学習の分野では、「教師データ」を用いて小規模な機械学習を行うことでモデル関数を構築し、そのモデル関数を用いて未知の変異体の活性を予測する方法が一般的である。そのための代表的な手法であるクロスバリデーション(CV)について以下に説明する (Figure 3-4-1)。

CVでは、「教師データ」を「学習用データ」と「精度検証用データ」に分割する。そして、「学習用データ」を用いてモデル関数の構築を行い、「精度検証用データ」を用いて、そのモデル関数を評価する方法である。例えば、「教師データ」を5分割し、そのうち4つを「学習用データ」としてモデル関数を構築し、残りの1つを「精度検証用データ」としてモデルの精度を評価する。さらに、「精度検証用データ」に用いるものを変えることにより、上記の検証を5回行い、その平均を取ることで精度を評価する。

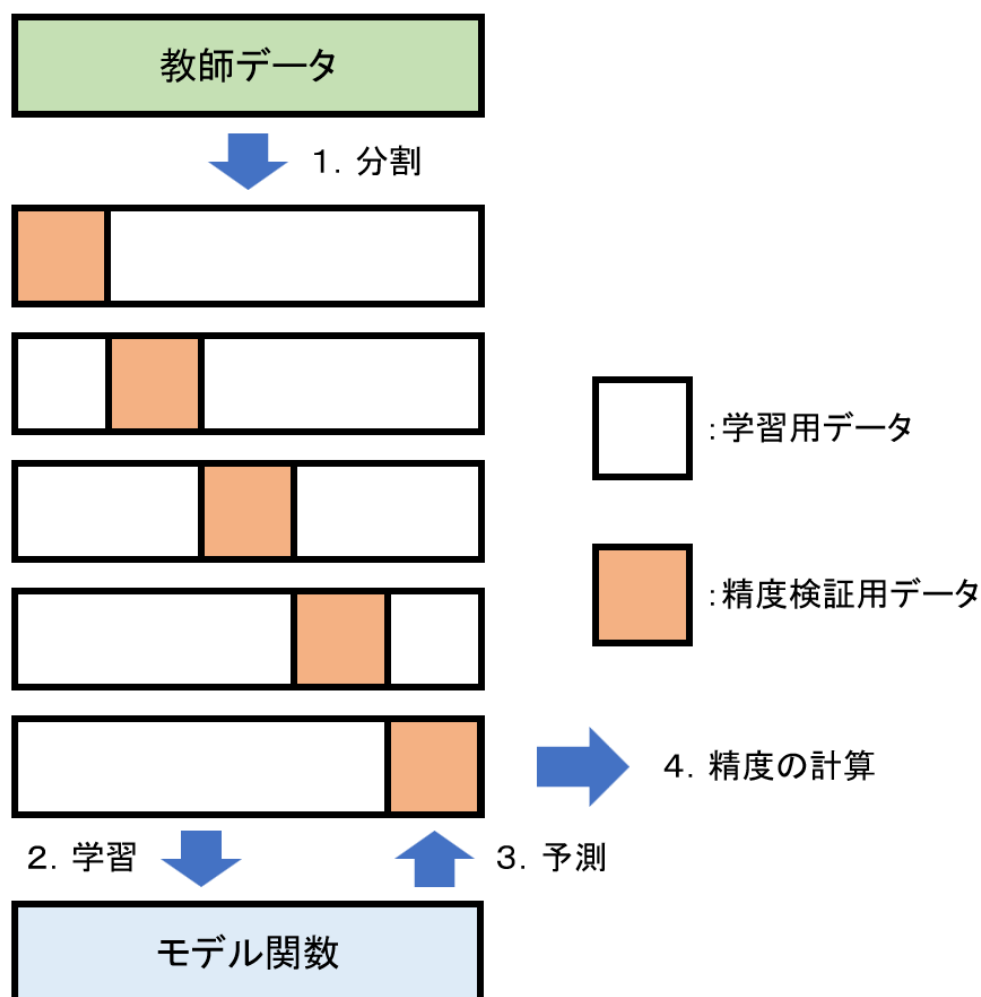


Figure 3-4-1. CVによるモデルの精度評価の概要。

上記の CV をさらに発展させた方法が **Nested CV** である。この方法では、通常の CV と同様に「教師データ」を「学習用データ」と「精度検証用データ」に分割する。「学習用データ」をさらに分割して小規模な CV を行い、モデル関数を修正することで、より正確な精度評価が可能になる。

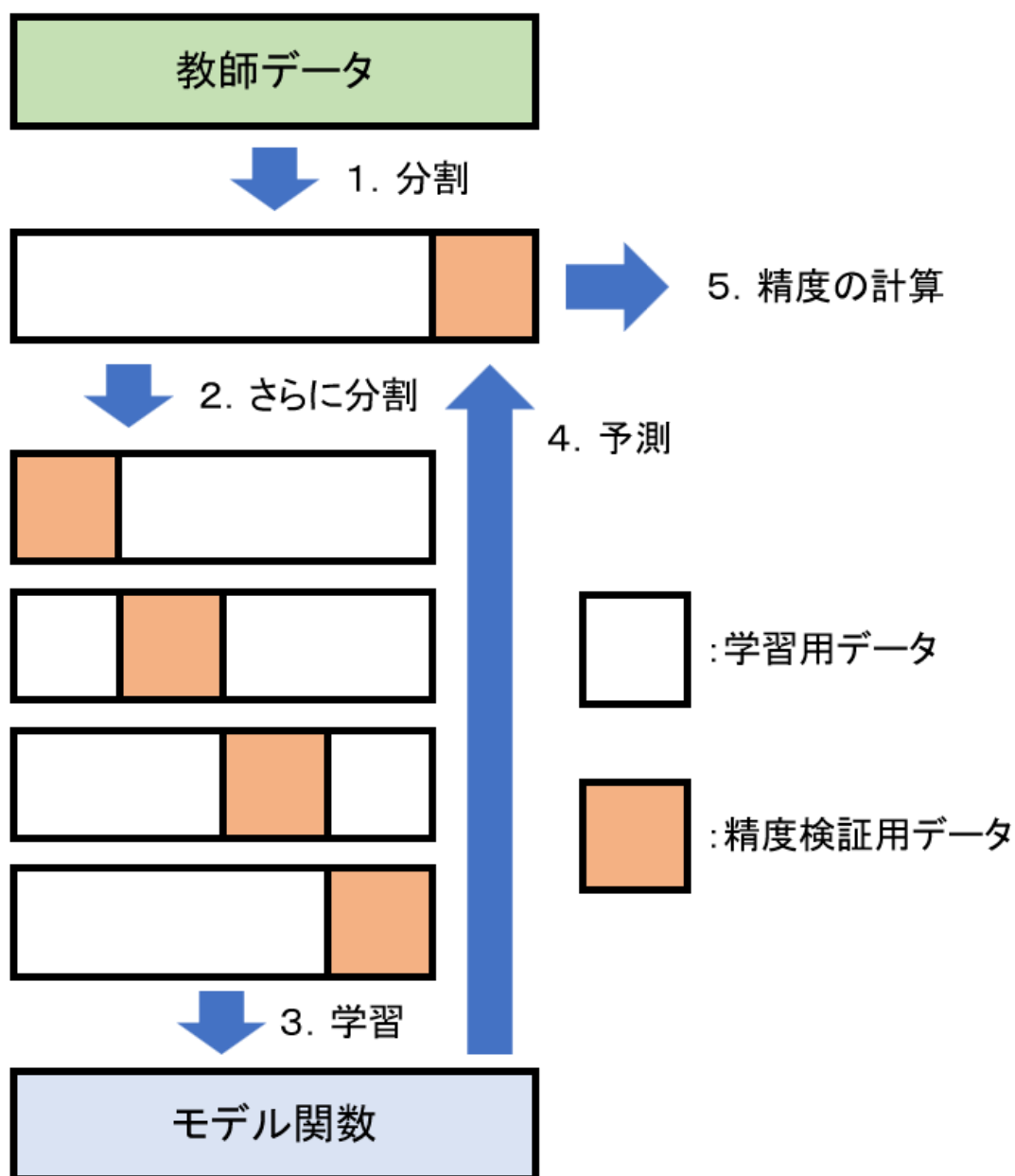


Figure 3-4-2. Nested CV によるモデルの精度評価の概要。



## 第五節 機械学習のモデル関数の構築

第二章の結果から NADPH の消費量と生成物の量の間には大きな乖離があることが判明したため、生成物の量を教師データとして機械学習に着手した。なお、変異体酵素（合計 164 種類）のうち不安定で精製できなかったもの（合計 62 種類）の生成物の量をゼロと定義することで、不安定で精製できない変異体を除外することを試みた。機械学習のモデルとしてはサポートベクター回帰<sup>13, 14)</sup>を用い、上述のように活性部位の 4 アミノ酸残基を Protein descriptor (ST-scale, Z-scale, T-scale, FASGAI, MS-WHIM, ProtFP, VHSE, BLOSUM, PSSM) や BERT を用いて特徴ベクトルに変換した。

モデルの精度評価をするために、Nested 5 fold CV を用いて、スピアマンの順位相関係数にてモデルの精度評価（Figure 3-5-1a）をした。BLOSUM と PSSM を組み合わせた特徴量（denoted as BLOSUM x PSSM）が最も良い精度を示すことが判明した。TAPE transformer 及び PHBH のホモログで fine-tuned された TAPE transformer（denoted as Evtuned-BERT）を用いて、同様に Nested-CV にて精度評価を行った。すると、Evtuned-BERT の方が BERT よりも良い予測精度を示し、Evtuned-BERT は BLOSUM x PSSM に匹敵する程度の精度（Figure 3-5-1b）を示した。

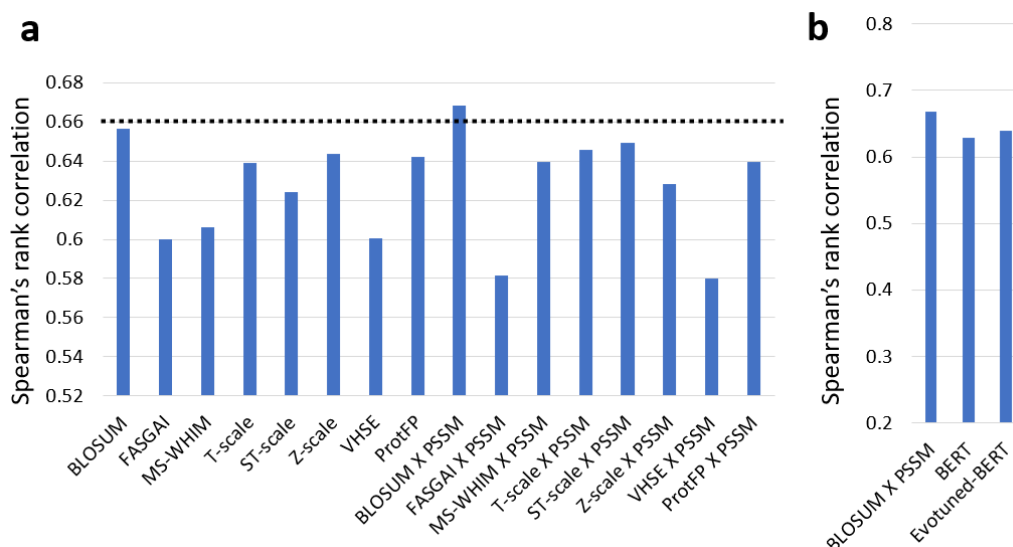
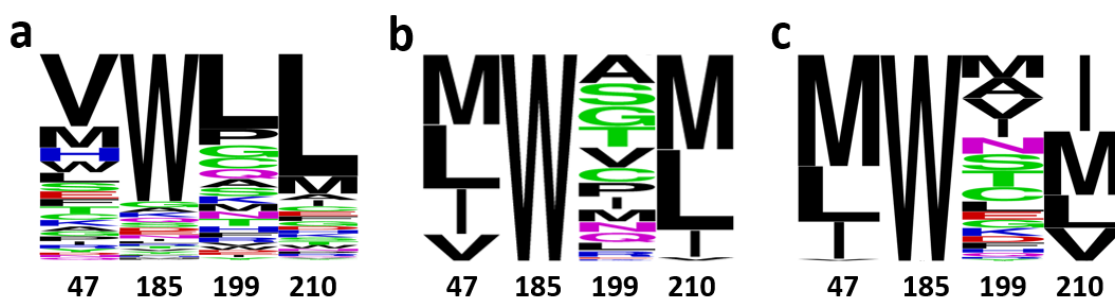


Figure 3-5-1. Nested 5 fold CV によるモデルの精度評価。(a) Protein descriptor 及び (b) BERT, Evtuned-BERT を用いたモデルの精度評価。

## 第六節 モデル関数を用いた全変異体の酵素活性の予想

BLOSUM x PSSM と Evtuned-BERT により構築された 2 つの機械学習のモデルを用いて全配列空間への予測をした。それぞれの上位 100 を抽出すると、両者の上位 100 変異体の配列は似ているが、ところどころ異なるような結果となった (Figure 3-6-1)。両者の 100 変異体中で 23 種類の変異体は共通していることが確認できた。共通している 23 種類の発現ベクターの作製を試みたところ、その中の 19 種類 (Table 3-6-1a) を作製することに成功した。その 19 種類の配列が (M/L)-W-X-(M/L) であり、Leu199 の部分は配列の多様性が出ているが、V47, W185, L210 については配列の多様性が乏しいことが判明した。

配列の多様性を出すことで、幅広い変異体酵素を検証すべく、それぞれの上位 100 種類の変異体の中から、手動で追加の変異体を選択した。初期のライブラリーでは Trp185 に変異が入った変異体は不安定であり、ほとんどが精製できなかった。また、一部の精製できた変異体 (PHBH-W185H, PHBH-MLPE) は活性が消失していたので、Trp185 の変異は活性の増加に繋がらないことが予想された。よって、配列の多様性を出すために配列が (M/L)-W-X-(M/L) で表わされず、かつ、Val47 と Leu210 に変異が入ったものを活性試験にて検証することにした。両者の上位 100 変異体を確認し、(I/K)-W-X-(M/L), L-W-X-(M/L), M-W-X-(I/V/K) で表される変異体を中心にして発現ベクターを作製したところ、25 種類の発現ベクター (Table 3-6-1b) の作製に成功した。合計 44 種類の変異体について、活性試験にて検証することにした。



**Figure 3-6-1.** シーケンスロゴによる配列の分析 (a) 初期ライブラリー (b) BLOSUM x PSSM model による上位 100 変異体 (c) Evtuned-BERT model による上位 100 変異体

**Table 3-6-1.** セカンドライブラリーの配列と予測順位 **(a)** 両者のモデルにて上位 100 以内であると予想された変異体 **(b)** 片方のモデルにて、上位 100 以内であると予想された変異体

**a**

Variants	BLOSUM x PSSM	EvoTuned-BERT
LWMM	42	88
LWVL	16	87
MWAL	3	8
MWAM	1	29
MWCM	35	10
MWIL	41	7
MWIM	30	15
MWKM	99	26
MWLM	53	20
MWML	43	1
MWNL	69	16
MWQM	39	17
MWRL	93	39
MWRM	76	81
MWSL	7	25
MWVL	12	3
MWVM	8	5
VWCM	100	94
VWVM	50	59

**b**

Variants	BLOSUM x PSSM	EvoTuned-BERT
IWAI	65	13129
IWNM	97	1217
KWAL	176	82
KWCM	416	75
KWML	645	12
KWMM	484	13
KWVL	302	41
KWVM	253	46
LWAI	34	22556
LWAV	88	39245
LWGI	91	38422
LWNL	79	473
LWNM	62	1403
LWSI	64	41150
LWTI	85	54974
LWVI	75	5854
MWAI	28	636
MWAV	78	1254
MWGI	84	1125
MWMK	1082	92
MWMV	194	89
MWSI	52	1339
MWTI	74	2063
MWVI	66	142
VWAI	83	3584

## 第七節 機械学習により予測された高活性変異体の活性試験による

### 検証

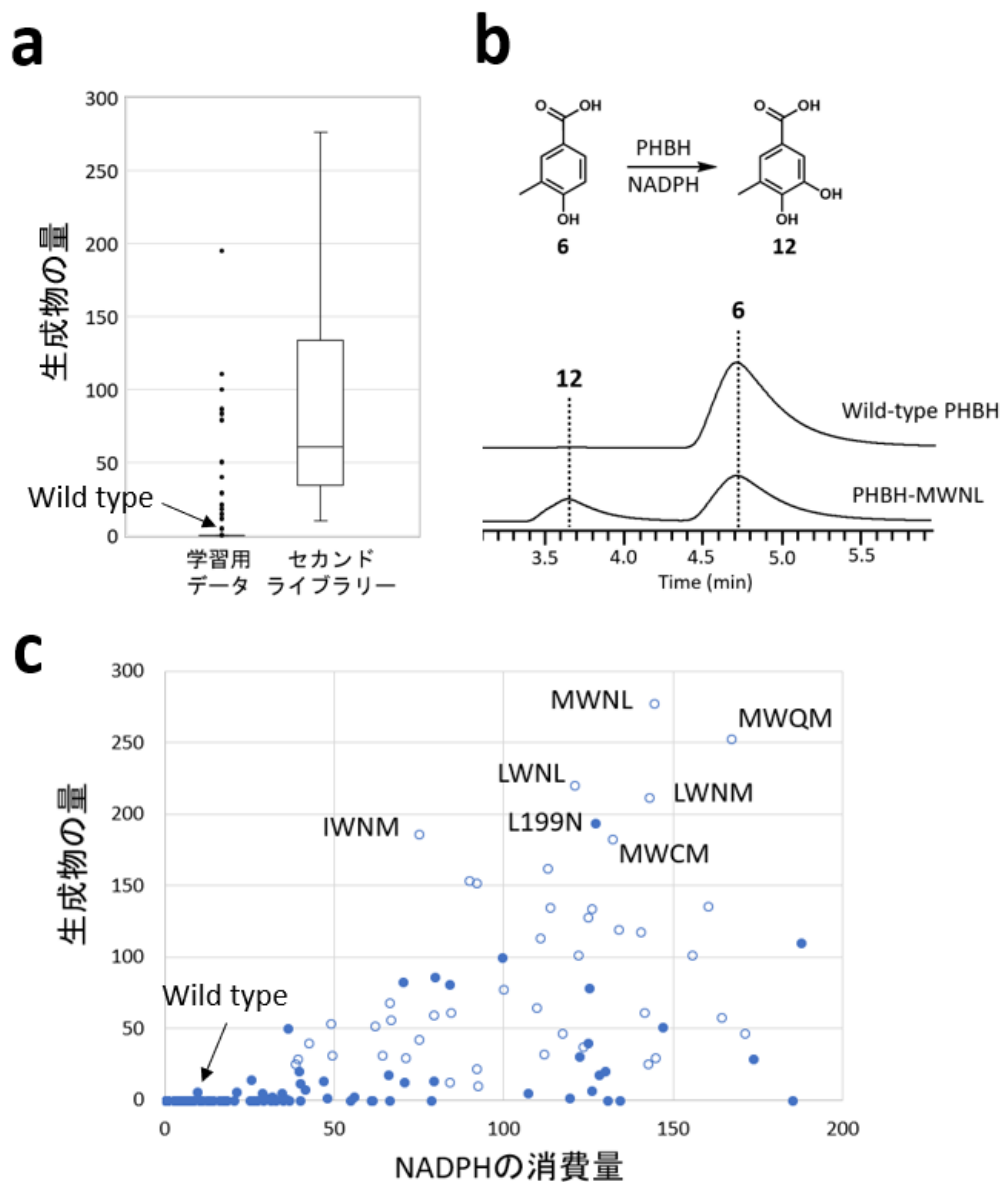
合計 44 種類の変異体について、活性試験による検証をすることにした。第二章と同様の方法にて、NADPH の消費量と生成物の収量を定量した。すると、初期のライブラリーでは変異体 164 種類の中で野生型よりも活性が高かった変異体は僅か 25 種類であった。それに対して、セカンドライブラリーでは、44 種類を実験的に検証したところ、44 種類全ての変異体が野生型酵素よりも活性が高い (Figure 3-7-1a, c) ことが判明した。また、その中で活性が最も強かった PHBH-MWNL は野生型酵素よりも活性が大幅に上昇していた (Figure 3-7-1b)。

このことから、機械学習により予想されたセカンドライブラリーは活性が高い変異体酵素が濃縮されていることが判明した。セカンドライブラリーの中の上位 4 変異体 (活性が高い順に、PHBH-MWNL, PHBH-MWQM, PHBH-LWNL, PHBH-LWNL) は、初期のライブラリーで最も活性が高かった PHBH-L199N よりも、さらに活性が高いことが判明した (Figure 3-7-1c)。さらに、PHBH-IWNM (セカンドライブラリーで 5 位), PHBH-MWCM (セカンドライブラリーで 6 位) は PHBH-L199N に匹敵するほどの活性を有していることが判明した (Figure 3-7-1c)。PHBH-MWCM, PHBH-MWNL, PHBH-MWQM は BLOSUM x PSSM と Evtuned-BERT の両者にて上位 100 以内 (Figure 3-6-1a) であると予想された。それに対して、PHBH-IWNM, PHBH-LWNL, PHBH-LWNL は BLOSUM x PSSM を用いたモデルのみで上位 100 以内 (Figure 3-6-1b) であると予想された。つまり、複数のモデルを使用することにより 1 つのモデルだけでは見落としてしまうような変異体を発見できることも示唆された。

セカンドライブラリーで最も活性が強かった 6 種類の変異体 (PHBH-MWNL, PHBH-MWQM, PHBH-LWNL, PHBH-LWNL, PHBH-IWNM, PHBH-MWCM) の中で、PHBH-MWNL, PHBH-MWQM, PHBH-MWCM の活性が強いは予想することができた。なぜなら、PHBH-V47M, PHBH-L199C/N/Q や PHBH-L210M も活性が強かったためである。それに対して、PHBH-V47L や PHBH-V47I が活性をほとんど持っていないにもかかわらず、PHBH-LWNL, PHBH-LWNL, PHBH-IWNM が強い活性を持っていたことから、一重変異体では活性が弱くても組み合わせることで高活性変異体を創出できることが示唆された。

また、初期のライブラリーでは、NADPH がほとんど消費されない変異体が多数存在し、NADPH が消費された変異体の中でも、生成物がほとんど生産されない変異体が多数確認された。しかし、セカンドライブラリーでは全ての変

異体で生成物を確認できたので、機械学習は FMO によるアンカップリングの改善にも有効であることが示された。



**Figure 3-7-1.** セカンドライブラリーの活性試験 (a) 初期のライブラリーとセカンドライブラリーの生成物量の箱ひげ図による比較 (b) Wild-type PHBH と PHBH-MWNL の生成物量の比較 (260 nm の紫外線吸収で測定している) (c) NADPH 消費量と生成物量の比較。塗りつぶし：初期ライブラリー (165 種類)、塗りつぶし無し：セカンドライブラリー (44 種類)

## 第八節 PHBH-MWNL の速度論的な解析

機械学習により発見された 44 種類の高活性変異体の中で、PHBH-MWNL が最も多くの生成物を与えた。その詳細な機能について考察をすべく、ミカエリス・メンテン式を用いた速度論的な解析に着手した。すると、wild-type PHBH は  $K_M = 472.2 \pm 70.7 \mu\text{M}$ 、 $k_{cat} = 0.36 \pm 0.02 \text{ min}^{-1}$ 、PHBH-MWNL は  $K_M = 178.8 \pm 18.6 \mu\text{M}$ 、 $k_{cat} = 56.3 \pm 1.9 \text{ min}^{-1}$ であった (Figure 3-8-1)。上記の結果より、PHBH-MWNL の **6** に対する  $K_M$  は、wild-type PHBH の半分以下であった。

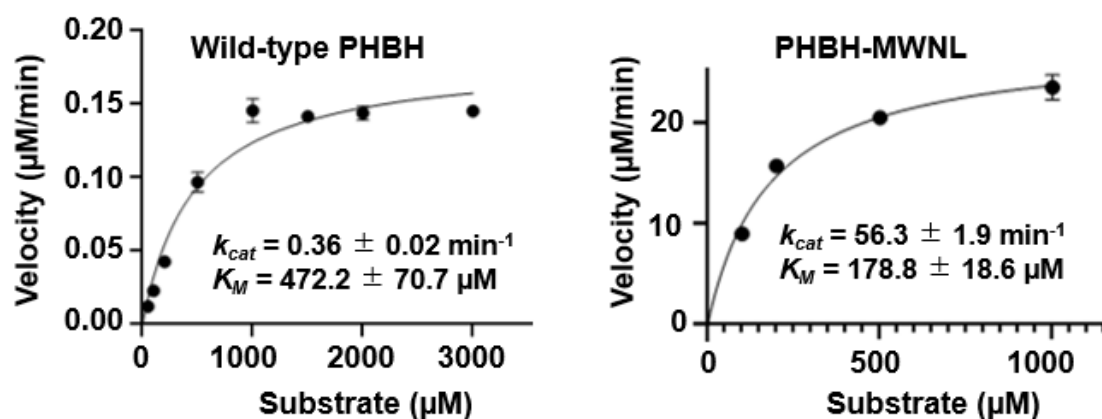


Figure 3-8-1. Wild-type PHBH 及び PHBH-MWNL のミカエリス・メンテン式を用いた速度論的な解析 (標本数は 3、エラーバーは標準偏差)

**1** と **6** を比較すると、**6** はメチル基を有しているため **1** よりも立体障害が大きい。よって、PHBH-MWNL では V47 が Met に置換されたことでポケットの構造が柔軟になり、嵩高い基質である **6** に対する親和性が上昇したことが予想される。

代謝回転数  $k_{cat}$  については、PHBH-MWNL の方が wild-type PHBH よりも 100 倍以上も上昇していた。初期のライブラリーでは PHBH-L199N や PHBH-L199Q などの L199 を側鎖にアミド基を持つアミノ酸に変換したものでは活性が上昇していた。さらには、**6** は疎水性が高い構造を有しているため、L199 を極性のアミノ酸残基に置換することで **6** に対する親和性が上昇するとは考えにくい。よって、L199 を Asn に置換したことが基質認識性とは別のメカニズムで活性の増強に寄与していることが予想される。

それに対して、初期のライブラリーでは、PHBH-L199D や PHBH-L199E などの L199 を酸性のアミノ酸残基に置換した場合は活性が消失した。つまりは、

「極性が高く」、「酸性ではない」アミノ酸残基に置換すると活性が上昇することが判明した。そのメカニズムは大変興味深く、次章で結晶構造解析を用いて詳細なメカニズムの解明を行った。

## 第九節 小括

本章では、合計  $20^4=160,000$  種類の PHBH 変異体の中から、僅か 165 種類を教師データとして機械学習をすることで、160,000 種類全ての変異体の酵素活性を予測した。上位であると予測された変異体の中から 44 種類を活性試験にて検証することで、PHBH の高活性変異体を多数発見することに成功した。また、機械学習により発見された高活性変異体の中では、PHBH-MWNL が最も活性が強く、野生型酵素の 100 倍以上の  $k_{cat}$  を有することが判明した。活性が上昇したメカニズムが大変興味深く、次章では結晶構造解析により詳細な解析に着手した。



## 第十節 実験項

### 3.1 変異型 PHBH のプラスミド作製

- ・第二章と同様の手順にて行った。

### 3.2 大腸菌を用いた形質転換、培養手順

- ・第二章と同様の手順にて行った。

### 3.3 PHBH の精製手順

- ・第二章と同様の手順にて行った。

### 3.4 PHBH の活性試験の手順

- ・第二章と同様の手順にて行った。

### 3.5 Wild-type PHBH 及び PHBH-MWNL の速度論的な解析

0.5  $\mu\text{M}$  enzyme, 1 mM NADPH, 50  $\mu\text{M}$ -3000  $\mu\text{M}$  **6**, 100 mM KPb (pH 7.4), 100 mM NaCl (50  $\mu\text{L}$ ) で 25 °C, 10 min 活性試験した。1 N HCl を 5  $\mu\text{L}$  加えて pH を下げた後、100  $\mu\text{L}$  の酢酸エチルを用いて 2 回抽出した。抽出後の有機層を減圧下で濃縮した後、内部標準物質として 0.10 mM の trihydroxy-benzoic acid を含んだ 50  $\mu\text{L}$  の DMF に溶解した。第二章と同様にして HPLC にて **12** を紫外線吸収にて定量した。誤差を評価するために、 $n=3$  で測定した。グラフパッドプリズムにて、 $K_M$  ( $\mu\text{M}$ ) 及び  $V_{\text{max}}$  ( $\mu\text{M}/\text{min}$ ) を算出した。 $V_{\text{max}}$  を酵素濃度で割ることにより  $k_{\text{cat}}$  ( $\text{min}^{-1}$ ) を算出した。

**Table 3-10-1. 発現用ベクターと SDS-PAGE のリスト**

No	Vector map	Vector name	SDS-PAGE (Figure number- Gel name and lane number)	Mutation
1	Figure S3-G1	pKW18756-IWAI	Figure S5-A1	V47I/L199A/L210I
2	Figure S3-G2	pKW18756-IWNM	Figure S5-A2	V47I/L199N/L210M
3	Figure S3-G3	pKW18756-KWAL	Figure S5-A3	V47K/L199A
4	Figure S3-G4	pKW18756-KWCM	Figure S5-A4	V47K/L199C/L210M
5	Figure S3-G5	pKW18756-KWML	Figure S5-A5	V47K/L199M
6	Figure S3-G6	pKW18756-KWMM	Figure S5-A6	V47K/L199M/L210M
7	Figure S3-G7	pKW18756-KWVL	Figure S5-A7	V47K/L199V
8	Figure S3-G8	pKW18756-KWVM	Figure S5-A8	V47K/L199V/L210M
9	Figure S3-G9	pKW18756-LWAI	Figure S5-A9	V47L/L199A/L210I
10	Figure S3-G10	pKW18756-LWAV	Figure S5-A10	V47L/L199A/L210V
11	Figure S3-G11	pKW18756-LWGI	Figure S5-A11	V47L/L199G/L210I
12	Figure S3-G12	pKW18756-LWMM	Figure S5-B1	V47L/L199M/L210M
13	Figure S3-G13	pKW18756-LWNL	Figure S5-B2	V47L/L199N
14	Figure S3-G14	pKW18756-LWNM	Figure S5-B3	V47L/L199N/L210M
15	Figure S3-G15	pKW18756-LWSI	Figure S5-B4	V47L/L199S/L210I
16	Figure S3-G16	pKW18756-LWTI	Figure S5-B5	V47L/L199T/L210I
17	Figure S3-G17	pKW18756-LWVI	Figure S5-B6	V47L/L199V/L210I
18	Figure S3-G18	pKW18756-LWVL	Figure S5-B7	V47L/L199V
19	Figure S3-G19	pKW18756-MWAI	Figure S5-B8	V47M/L199A/L210I
20	Figure S3-G20	pKW18756-MWAL	Figure S5-B9	V47M/L199A
21	Figure S3-G21	pKW18756-MWAM	Figure S5-B10	V47M/L199A/L210M
22	Figure S3-G22	pKW18756-MWAV	Figure S5-C1	V47M/L199A/L210V
23	Figure S3-G23	pKW18756-MWCM	Figure S5-C2	V47M/L199C/L210M
24	Figure S3-G24	pKW18756-MWGI	Figure S5-C3	V47M/L199G/L210I
25	Figure S3-G25	pKW18756-MWIL	Figure S5-C4	V47M/L199I
26	Figure S3-G26	pKW18756-MWIM	Figure S5-C5	V47M/L199I/L210M
27	Figure S3-G27	pKW18756-MWKM	Figure S5-C6	V47M/L199K/L210M
28	Figure S3-G28	pKW18756-MWLM	Figure S5-C7	V47M/L210M
29	Figure S3-G29	pKW18756-MWMK	Figure S5-C8	V47M/L199M/L210K
30	Figure S3-G30	pKW18756-MWML	Figure S5-C9	V47M/L199M
31	Figure S3-G31	pKW18756-MWMV	Figure S5-C10	V47M/L199M/L210V

32	Figure S3-G32	pKW18756-MWNL	Figure S5-C11	V47M/L199N
33	Figure S3-G33	pKW18756-MWQM	Figure S5-C12	V47M/L199Q/L210M
34	Figure S3-G34	pKW18756-MWRL	Figure S5-D1	V47M/L199R
35	Figure S3-G35	pKW18756-MWRM	Figure S5-D2	V47M/L199R/L210M
36	Figure S3-G36	pKW18756-MWSI	Figure S5-D3	V47M/L199S/L210I
37	Figure S3-G37	pKW18756-MWSL	Figure S5-D4	V47M/L199S
38	Figure S3-G38	pKW18756-MWTI	Figure S5-D5	V47M/L199T/L210I
39	Figure S3-G39	pKW18756-MWVI	Figure S5-D6	V47M/L199V/L210I
40	Figure S3-G40	pKW18756-MWVL	Figure S5-D7	V47M/L199V
41	Figure S3-G41	pKW18756-MWVM	Figure S5-D8	V47M/L199V/L210M
42	Figure S3-G42	pKW18756-VWAI	Figure S5-D9	L199A/L210I
43	Figure S3-G43	pKW18756-VWCM	Figure S5-D10	L199C/L210M
44	Figure S3-G44	pKW18756-VWVM	Figure S5-E1	L199V/L210M

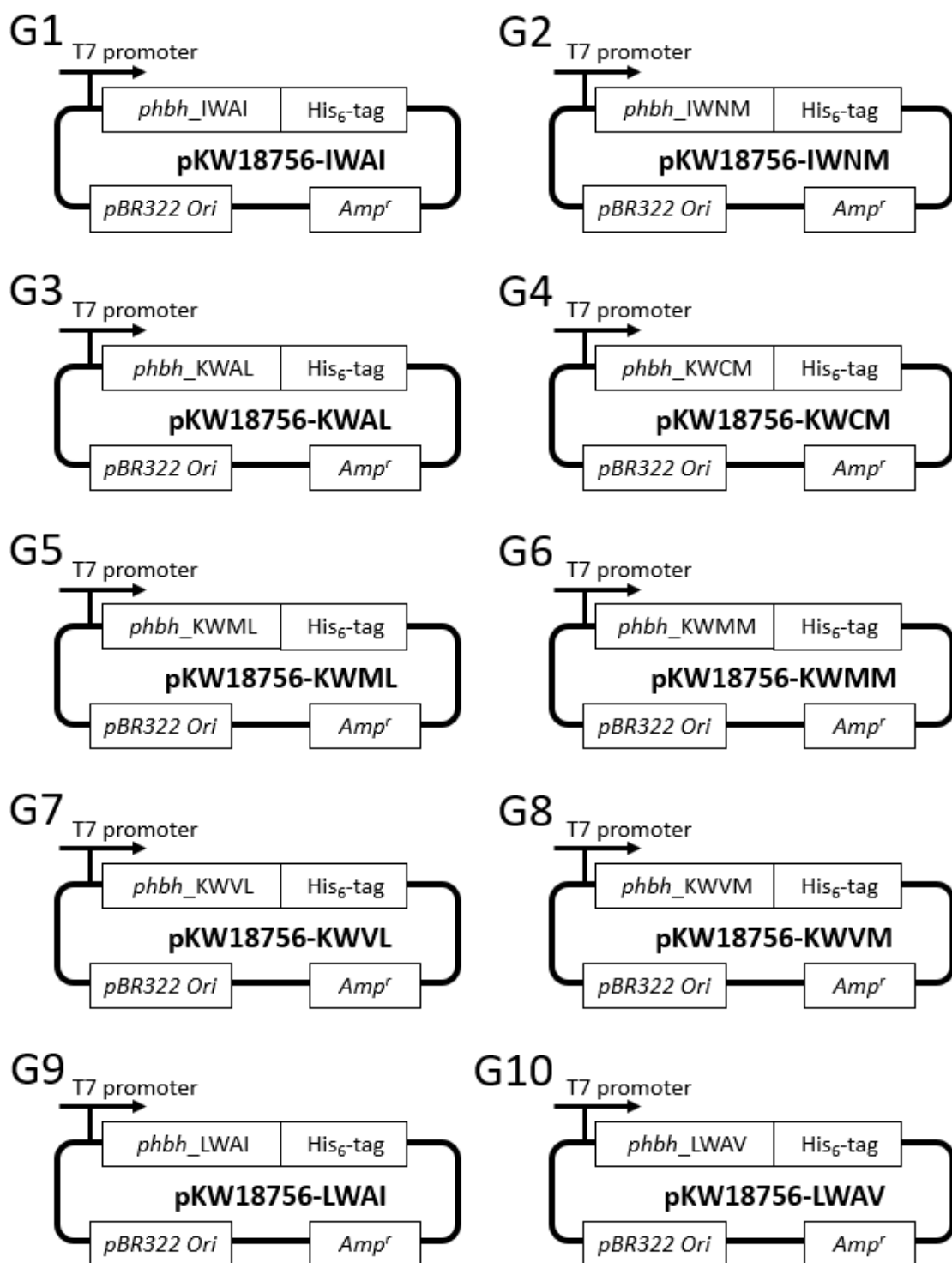


Figure 3-10-1. セカンドライブラリー作製のベクターマップ

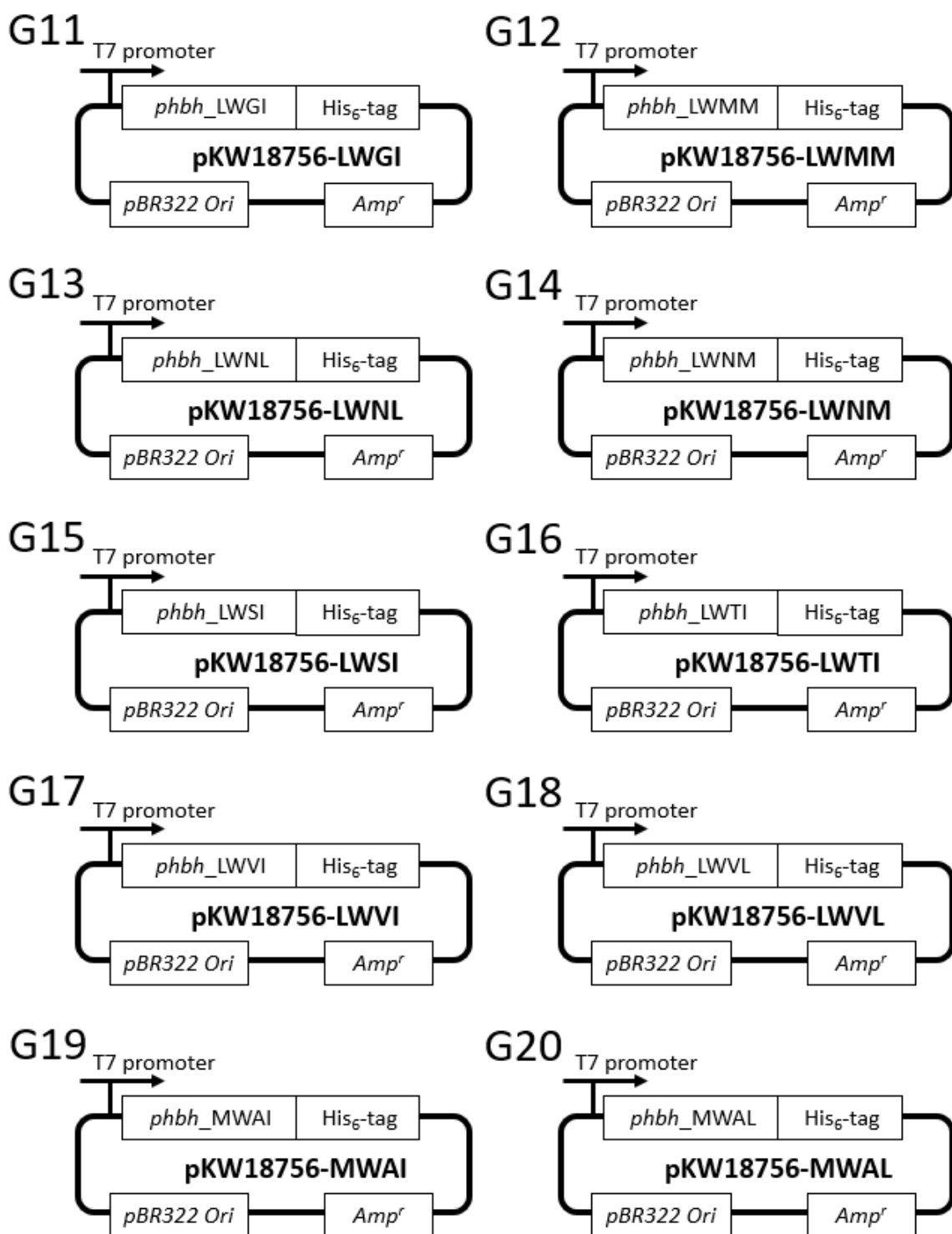


Figure 3-10-1. (前ページの続き)

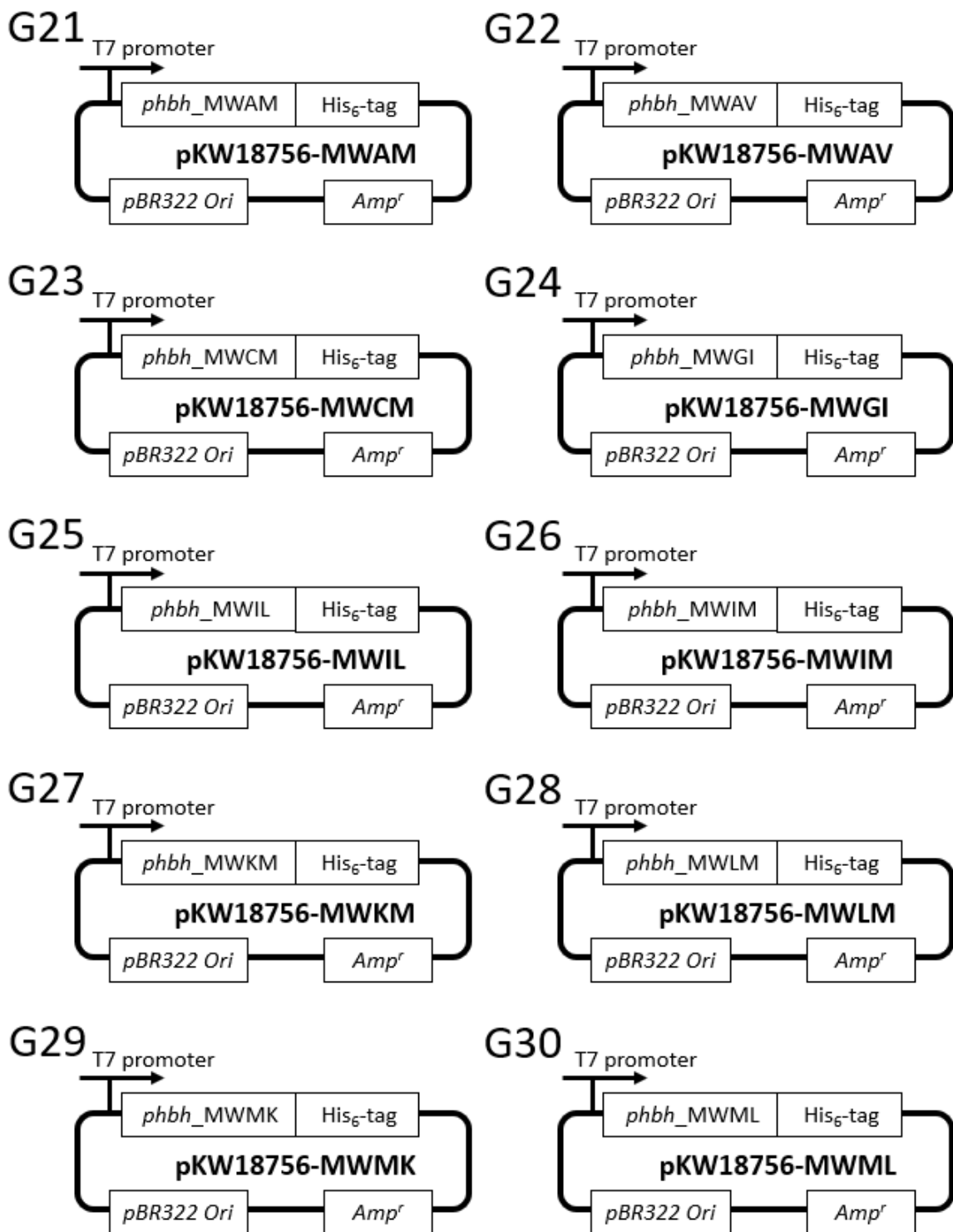


Figure 3-10-1. (前ページの続き)

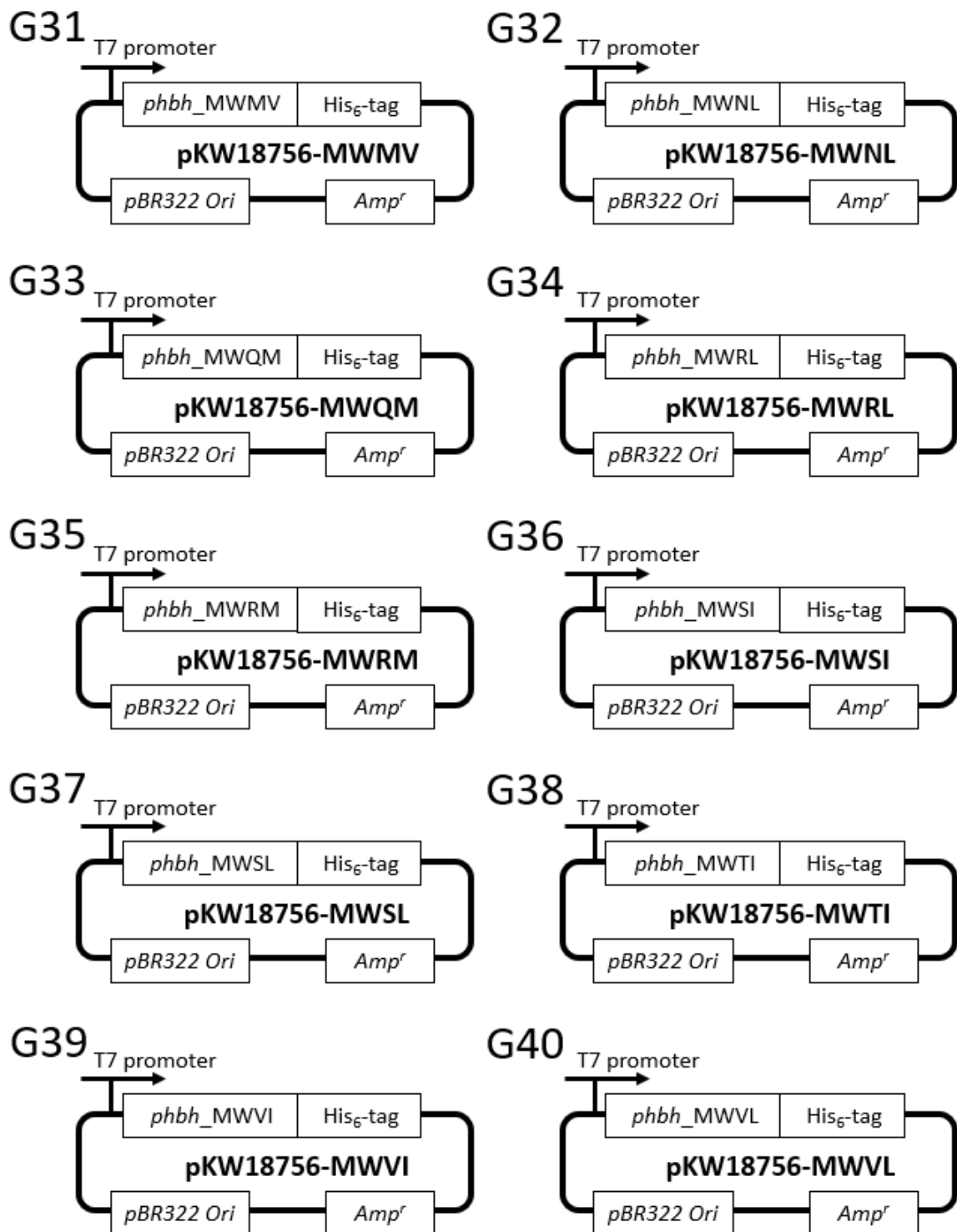


Figure 3-10-1. (前ページの続き)

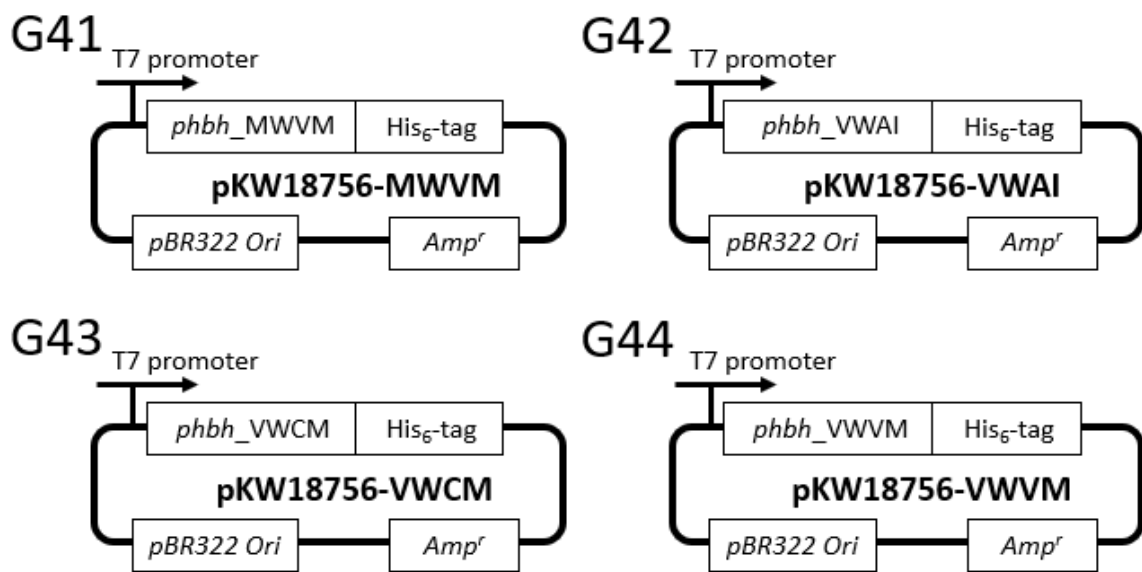
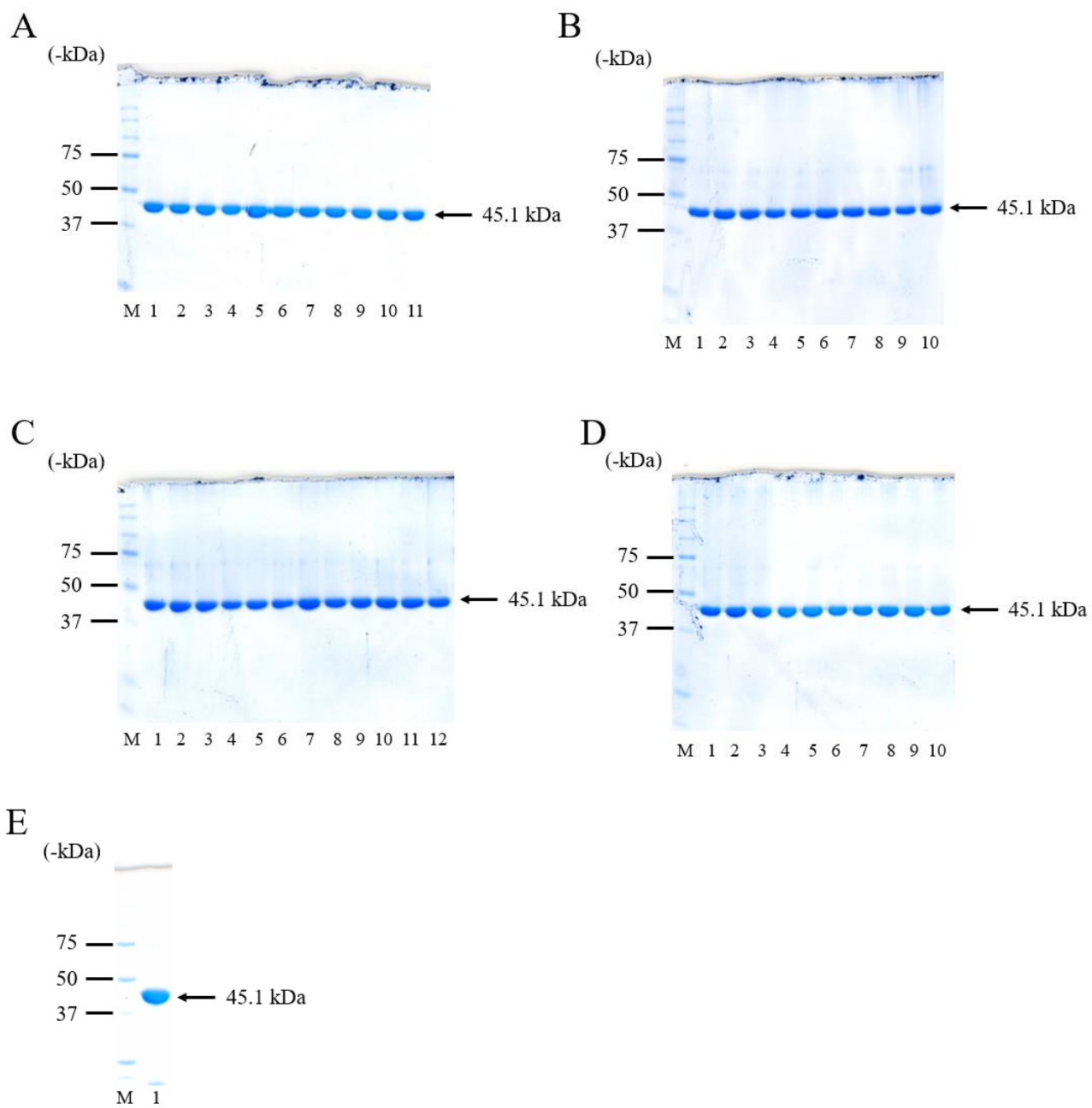


Figure 3-10-1. (前ページの続き)





**Figure 3-10-2.** セカンドライブラリーの SDS-PAGE

## 第十一節 引用文献

- (1) Yang, L.; Shu, M.; Ma, K.; Mei, H.; Jiang, Y.; Li, Z. ST-scale as a novel amino acid descriptor and its application in QSAM of peptides and analogues. *Amino Acids* **2010**, *38*, 805-816.
- (2) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J. Med. Chem.* **1998**, *41*, 2481-2491.
- (3) Tian, F.; Zhou, P.; Li, Z. T-scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides. *J. Mol. Struct.* **2007**, *830*, 106-115.
- (4) Liang, G.; Li, Z. Factor Analysis Scale of Generalized Amino Acid Information as the Source of a New Set of Descriptors for Elucidating the Structure and Activity Relationships of Cationic Antimicrobial Peptides. *QSAR Comb. Sci.* **2007**, *26*, 754-763.
- (5) Zaliani, A.; Gancia, E. MS-WHIM Scores for Amino Acids: A New 3D-Description for Peptide QSAR and QSPR Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 525-533.
- (6) van Westen, G. J.; Swier, R. F.; Cortes-Ciriano, I.; Wegner, J. K.; Overington, J. P.; Ijzerman, A. P.; van Vlijmen, H. W.; Bender, A. Benchmarking of protein descriptor sets in proteochemometric modeling (part 2): modeling performance of 13 amino acid descriptor sets. *J. Cheminform* **2013**, *5*, 42.
- (7) Mei, H.; Liao, Z. H.; Zhou, Y.; Li, S. Z. A new set of amino acid descriptors and its application in peptide QSARs. *Pept. Sci.* **2005**, *80*, 775-786.
- (8) Georgiev, A. G. Interpretable Numerical Descriptors of Amino Acid Space. *J. Comput. Biol.* **2009**, *16*, 703-723.
- (9) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.;

Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25*, 3389-3402.

(10) Rao, R.; Bhattacharya, N.; Thomas, N.; Duan, Y.; Chen, X.; Canny, J.; Abbeel, P.; Song, Y. S. Evaluating Protein Transfer Learning with TAPE. *Adv Neural Inf Process Syst* **2019**, *32*, 9689-9701.

(11) Rivesa, A.; Meiera, J.; Sercua, T.; Goyala, S.; Linb, Z.; Liua, J.; Guo, D.; Otta, M.; Zitnicka, C. L.; Mad, J.; Fergus, R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2016239118.

(12) Yamaguchi, H.; Saito, Y. Evotuning protocols for Transformer-based variant effect prediction on multi-domain proteins. *Briefings Bioinf.* **2021**, *22*, bbab234.

(13) Smola, A. J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199-222.

(14) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825-2830.

## 第四章 PHBH-MWNL の活性向上メカニズムの解明

### 第一節 PHBH-MWNL の結晶化

上記の件について詳細なメカニズムを解明するために、PHBH-MWNL の結晶構造解析に着手した。大スケールにて培養し、高純度なタンパク溶液を獲得することに成功した。収量は十分ではあったものの、PHBH-MWNL は高濃度では不安定であり、3 mg/mL 程度までしか濃縮することができなかった。通常タンパク質の結晶構造解析をする場合は、10-50 mg/mL 程度の高濃度のタンパク溶液が必要であるため、結晶化は厳しいように思われた。

2.6 mg/mL の PHBH-MWNL に終濃度が 0.50 mM の **6** を溶解し、シッティングドロップ法にて一次スクリーニングをし、見つかった結晶化条件をハンギングドロップ蒸気拡散法にて最適化することにより、タンパク質の結晶を獲得することに成功した。

上記の単結晶を用いて SLS にて回折実験をすることで、データ収集をした。分解能 1.95 Å にて回折データを獲得することに成功し、空間群が C222<sub>1</sub> であることが確認された。これまでに、PHBH-A45G,<sup>1)</sup> PHBH-C116S,<sup>2)</sup> PHBH-R220Q,<sup>3)</sup> PHBH-Y201F,<sup>4)</sup> PHBH-Y222F,<sup>5)</sup> PHBH-N300D,<sup>4)</sup> PHBH-Y385F<sup>4)</sup> などの様々な変異体 PHBH の結晶構造が解明されてきた。その中で、PHBH-R220Q のみはアポ体での結晶構造が解明されているが、他の PHBH は全て基質との共結晶構造が解明されている。例えば、野生型 PHBH では基質無し状態で結晶化を試みたが、結晶の品質が悪く回折データの収集ができなかった。よって、基質が結合していない条件では結晶化は困難であるため、共結晶構造の獲得は容易であると思われた。しかし、回折データをもとに解析をすると、予想外に **6** の電子密度を確認することはできなかった (Figure 4-1-1)。このことから、**6** との共結晶構造を解明するためには、更なる条件検討が必要である。

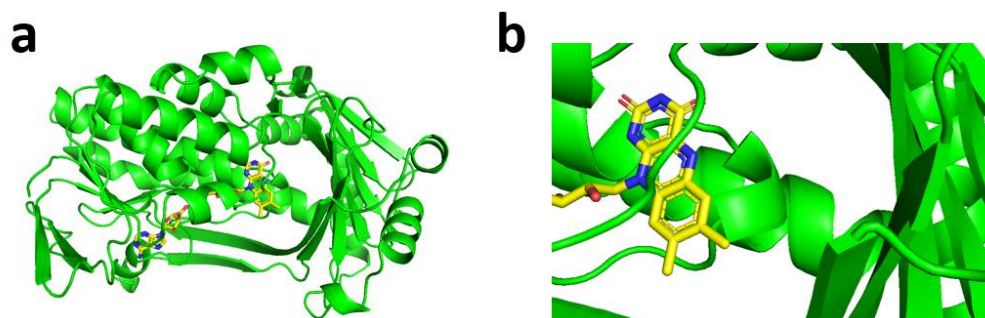


Figure 4-1-1. PHBH-MWNL の (a) 全体構造 (b) 基質結合部位の構造

## 第二節 PHBH-MWNL の共結晶化条件の検討

PHBH-MWNL-6 の共結晶構造を解明するために、より詳細な条件検討をすることにした。まず、異なる結晶化条件を探索することを考えた。結晶の空間群が変わることで、基質が活性部位に入りやすくなることがあるためである。しかし、これまでに解明された全ての結晶構造は空間群が  $C222_1$  であるため、別の空間群での結晶構造の獲得が困難であると予想された。

同一の空間群での基質との共結晶を獲得するために、より高濃度での結晶化に着手した。まず、タンパク溶液に高濃度の化合物を溶解することを試みたが、化合物の溶解度が低く 5 mM 程度の化合物を加えると溶け残ってしまった。そこで、リザーバーに化合物を溶解することを検討した。リザーバーには沈殿剤として有機化合物である PEG が溶解しているため、有機化合物との親和性が高く、より高濃度の化合物の溶解が可能になることが期待された。実際に条件検討をすると、80 mM 程度の 6 が溶解したリザーバーを作製することに成功した。

第一節と同様に、ハンギングドロップ蒸気拡散法にて条件検討をすることで単結晶を獲得し、高エネルギー加速器研究機構の PF にて X 線回折実験を行った。データ解析をすることにより、分解能 2.0 Å にて、PHBH-MWNL-FAD-6 の複合体の共結晶構造 (Figure 4-2-1) を解明することに成功した。

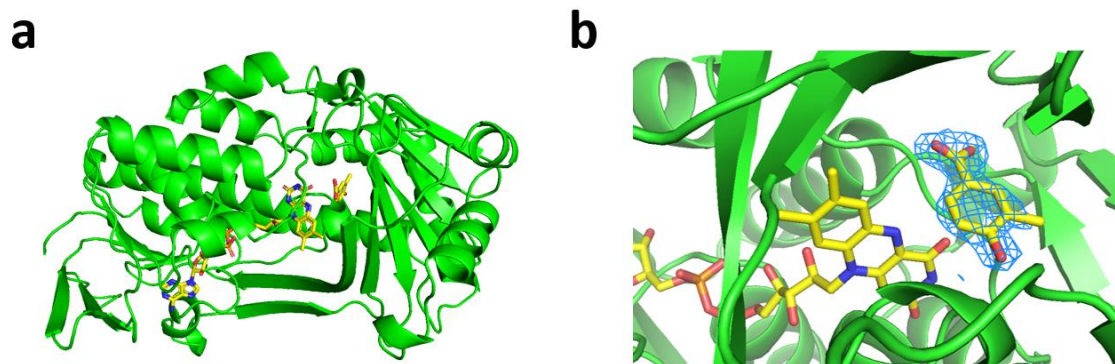


Figure 4-2-1. (a) PHBH-MWNL-FAD-6 複合体の全体構造 (b) FAD 近傍に確認できた 6 の電子密度 ( $2F_o-F_C$  map)

### 第三節 PHBH-MWNL の共結晶構造の解析

第二節で解明された PHBH-MWNL-FAD-6 複合体の結晶構造をもとにして、活性が大幅に上昇したメカニズムを解明することにした。まず、PHBH-MWNL の活性部位を確認すると、M47 の側鎖の硫黄原子と、N199 の側鎖のアミド基が水素結合を形成していることが確認された。それにより、各アミノ酸残基の向きが変わり、6 のメチル基が収納される空間が形成されることで、親和性が上昇していることが判明した。

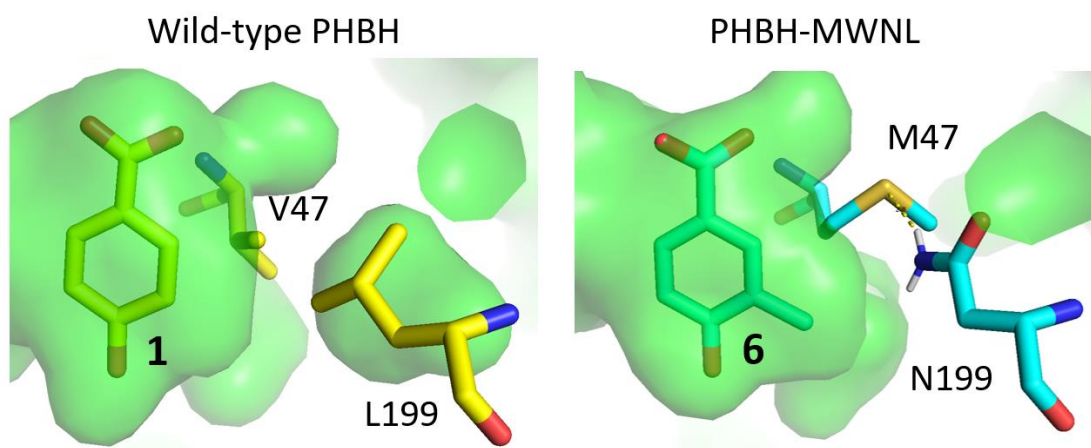
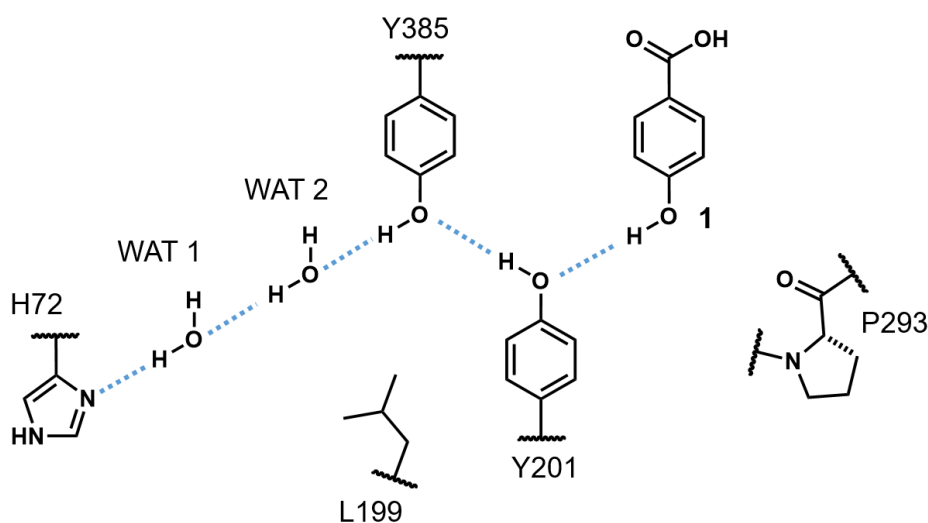


Figure 4-3-1. Wild-type PHBH 及び PHBH-MWNL の活性部位の比較

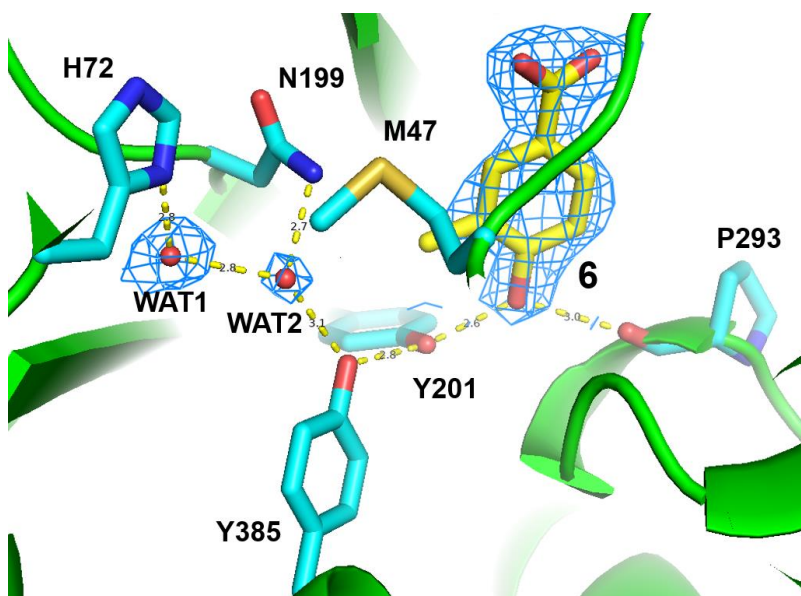
先行研究において、wild-type PHBH の反応メカニズムは以下のように提唱されている。まず初めに、His72, WAT1, WAT2, Tyr385, Tyr201、1 の順にプロトンリレーが形成されることで、基質である 1 の水酸基が脱プロトン化され、フェノラートイオンが形成する (Figure 4-3-2)。そして、形成されたフェノラートイオンと P293 の主鎖のカルボニル基との静電的な反発が原動力となり、酵素反応が進行することが予想されている<sup>6)</sup>。しかし、先行研究にて解明された wild-type PHBH の結晶構造では、存在するはずの WAT2 の電子密度が確認できなかった。WAT2 の周りに極性のアミノ酸残基が存在しないため、WAT2 を特定の位置に保持できないことが原因と考えられた。





**Figure 4-3-2.** プロトンリレーによる **1** のフェノール性水酸基の脱プロトン化機構

それに対して、PHBH-MWNL では N199 の周辺に、野生型 PHBH では確認できなかった WAT2 の電子密度を確認することに成功した (**Figure 4-3-3**)。PHBH-MWNL では L199 が Asn に置換されており、Asn の側鎖のアミド基が WAT2 と水素結合を形成することができる。それにより、プロトンリレーが安定化された状態になることが、 $k_{cat}$  の大幅な上昇に繋がったと考えられる。



**Figure 4-3-3.** 基質の脱プロトン化のためのプロトンネットワーク (電子密度は  $2F_o - F_c$  map)

#### 第四節 小括

今回の PHBH-MWNL-6 の結晶構造解析により、酵素活性が上昇しているメカニズムを解明することに成功した。また、従来の酵素の機能改変では、基質認識性に注目しがちであった。しかし、プロトンリレーの安定化により、酵素活性が飛躍的に上昇したことは大変興味深く、酵素の機能改変の新たな手法になる可能性が示唆された。



## 第五節 実験項

### 4.1 大腸菌を用いた形質転換、培養手順

pKW18756-MWNL を 0.5  $\mu\text{L}$  を取り、大腸菌 BL21(DE3)株に加え、1.5 kV でエレクトロポレーションした。これを LB/Amp 平板培地に塗布し、37  $^{\circ}\text{C}$  で一晩培養した。生えてきたコロニーを取り、2 mL の LB/Amp 液体培地に植菌し、37  $^{\circ}\text{C}$ , 180 rpm で一晩振とう培養した。前培養液を 1.0 mL ずつ取り、1 L x 5 の LB/Amp 液体培地にそれぞれ植菌し、37  $^{\circ}\text{C}$ , 180 rpm で振とう培養した。OD<sub>600</sub> が 0.5-0.6 に到達後、十分に氷冷した後、終濃度が 0.1 mM になるように IPTG を加え、18  $^{\circ}\text{C}$ , 180 rpm, 20 hours 振とう培養した。その後、培養液を 500 mL の遠心管に加え、8,000 rpm, 4  $^{\circ}\text{C}$ , 5 min で遠心分離した。菌体を水に懸濁し、50 mL のファルコンチューブに移し、2,500 G, 4  $^{\circ}\text{C}$ , 10 min で遠心分離した。上清を捨て、-80  $^{\circ}\text{C}$  で保管した。

### 4.2 PHBH-MWNL の精製手順

集菌した大腸菌の菌体に破碎用緩衝液 (100 mM Tris-HCl pH 7.4 (4  $^{\circ}\text{C}$ ), 100 mM NaCl, 50  $\mu\text{M}$  FAD) を 30 mL 加えて、大腸菌を懸濁した。フレンチプレス (100 MPa, twice) にて破碎し、10,000 G, 4  $^{\circ}\text{C}$ , 20 min で遠心分離した。上清を別の 50 mL チューブに移し、1 mL の Ni-NTA agarose を加え、4  $^{\circ}\text{C}$  で一時間以上転倒混和した。Ni-NTA agarose が黄色く着色していることが確認できた。混合液をカラムに通した後、20 mM imidazole を含む破碎用緩衝液を 30 mL 用いて、樹脂をリンスした後、100, 200 mM imidazole を含む緩衝液をそれぞれ 6.0 mL 用いてタンパク質を溶出した。すると、Ni-NTA agarose は元の色に戻り、タンパク溶液をは濃い黄色に着色している事が確認できた。SDS-PAGE (Figure 4-5-1a) にて、目的の分子量のバンドを確認できた。

その後、緩衝液として 10 mM Tris-HCl pH 7.4 (4  $^{\circ}\text{C}$ ), 100 mM NaCl を用いて、スーパーデックス 75、流速 1.0 mL/min でゲル濾過クロマトグラフィー (Figure 4-5-2) にて精製した。280 nm の紫外線吸収が確認できたフラクションを取り、SDS-PAGE (Figure 4-5-1b) にて確認をした。すると、タンパク質が十分な純度にて精製できたことを確認した。ブラッドフォード法にてタンパク質の濃度を推定すると、3.2 mg/mL であった。構造安定化のため終濃度が 0.50 mM になるように FAD を加えた後、100  $\mu\text{L}$  ずつ分注し液体窒素にて急速冷凍した。-80  $^{\circ}\text{C}$  で保管した。

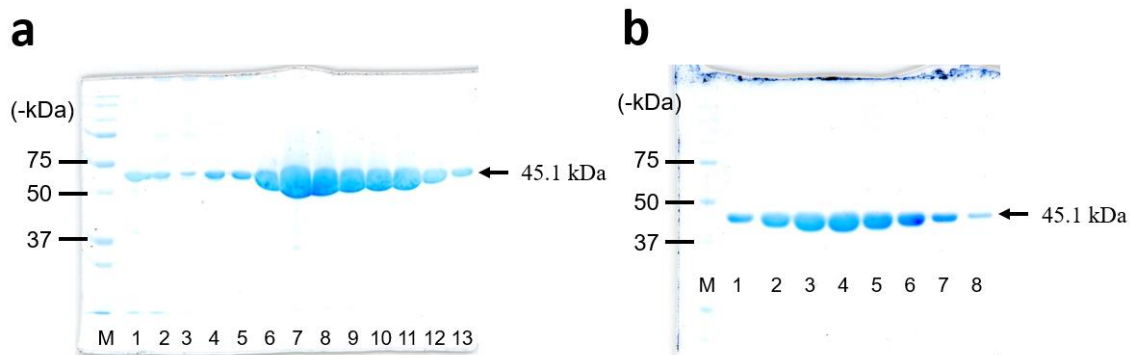
### 4.3 結晶化スクリーニング

PHBH-MWNL の結晶化条件を探索するために、シッティングドロップ蒸気拡散法 (**Table 4-5-1**) にて、結晶化スクリーニングを行った。すると、0.20 M sodium citrate, 20% PEGmme 3,350, 0.10 M Bis-trispropane (pH6.5) の条件において、単結晶 (**Figure 4-5-3a**) を獲得することに成功した。ハンギングドロップ蒸気拡散法 (**Table 4-5-2**) にて、塩濃度や沈殿剤の濃度を最適化することで、良質な単結晶 (**Figure 4-5-3b**) を獲得した。更に、高濃度の基質誘導体存在下で結晶を獲得すべく、リザーバーに 4-hydroxy-3-methylbenzoic acid (**6**) を溶解した条件で、ハンギングドロップ蒸気拡散法 (**Table 4-5-3**) に着手した。同様に 20 °C で3日間静置することで、単結晶 (**Figure 4-5-3c**) を獲得した。

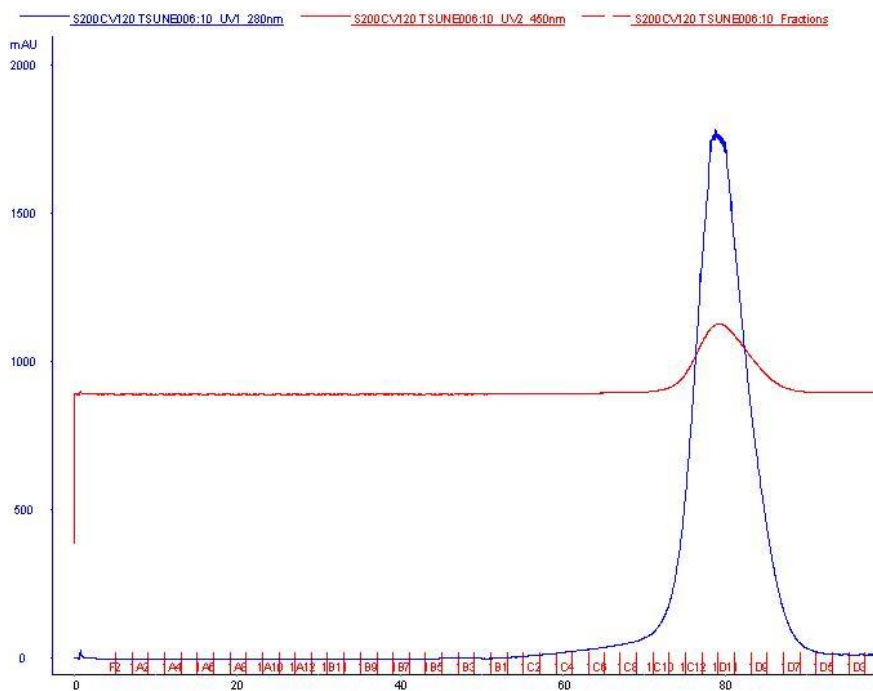
### 4.4 エックス線回折実験及びデータ解析

リザーバーに終濃度が 20% (v/v) になるようにエチレングリコールを加え、プロテクタントを作製した。プロテクタントにタンパク質結晶を数秒間浸して凍結処理を行った後、液体窒素を用いて瞬間凍結した。凍結した結晶は UniPuck (Crystal Positioning Systems) に充填し、液体窒素中にて保管した。高エネルギー加速器研究機構のビームライン BL17A を用いて、全自動測定にて X線回折実験 (**Table 4-5-4**) を行った。

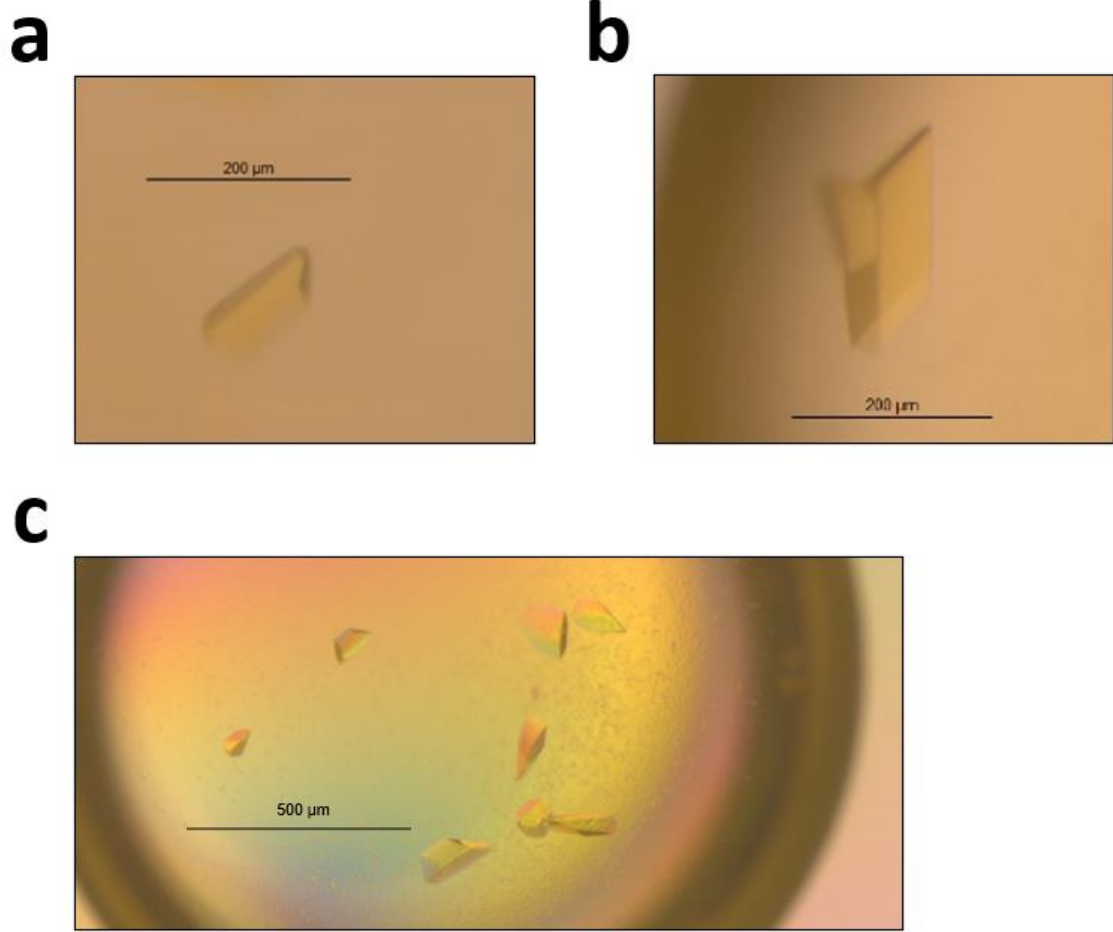
得られた X 線回折像をプログラム XDS<sup>7)</sup> 及び AIMLESS<sup>8)</sup>を用いて指数付け・積分処理し、回折強度データを得たところ、分解能 2.0 Å, 空間群 C222<sub>1</sub>, a=72.2, b=141.9, c=89.5 であることが確認できた。PHBH と生成物の複合体構造の結晶構造である 1PHH<sup>9)</sup>をサーチモデルとして、プログラム Phaser<sup>10)</sup>を用いて分子置換法を行った。プログラム Coot<sup>11)</sup>を用いてモデル修正を行い、プログラム Phenix.Refine<sup>12)</sup>により構造精密化を行った。**Table 4-5-5** に示すような統計値が算出された。



**Figure 4-5-1.** SDS-PAGE による PHBH-MWNL の分析 (a) M : マーカー、1:沈殿、2 : 上清、3 : フロースルー、4, 5 : 20 mM イミダゾールでリンスした画分、6, 7, 8, 9 : 100 mM イミダゾールで溶出した画分、9, 10, 11, 12, 13 : 200 mM イミダゾールで溶出した画分 (b) M : マーカー、1 : C11、2 : C12、3 : D12、4 : D11、5 : D10、6 : D9、7 : D8、8 : D7



**Figure 4-5-2.** ゲルろ過クロマトグラフィー



**Figure 4-5-3.** 獲得できた PHBH-MWNL の結晶 (a) シッティングドロップ蒸気拡散法で獲得できた結晶 (b) ハンギングドロップ蒸気拡散法で獲得できた結晶 (c) 80 mM の **6** 存在下で、ハンギングドロップ蒸気拡散法にて獲得できた結晶

**Table 4-5-1.** シットティングドロップによる条件検討

結晶化方法	シットティングドロップ蒸気拡散法
タンパク質溶液	2.6 mg/mL PHBH-MWNL, 0.50 mM 4-hydroxy-3-methylbenzoic acid ( <b>6</b> ), 0.50 mM FAD, 10 mM Tris-HCl (pH 7.4, 4 °C), 100 mM NaCl, 2% DMF
結晶化溶液	PACT suite (QIAGEN 社) (96 条件) The PEGs suite (QIAGEN 社) (96 条件) JCSG+ suite (QIAGEN 社) (96 条件) Crystal Screen I (HAMPTON RESEACH 社) (48 条件) Crystal Screen II (HAMPTON RESEACH 社) (48 条件) Wizard: Precipitant Synergy (EMERALD BIO 社) (192 条件)
混合比	リザーバー : タンパク質 = 0.50 $\mu$ L : 0.50 $\mu$ L
結晶化の温度	20 °C

**Table 4-5-2.** ハンギングドロップ蒸気拡散法による条件検討 1

結晶化方法	ハンギングドロップ蒸気拡散法
タンパク質溶液	2.6 mg/mL PHBH-MWNL, 0.50 mM 4-hydroxy-3-methylbenzoic acid ( <b>6</b> ), 0.50 mM FAD, 10 mM Tris-HCl (pH 7.4, 4 °C). 100 mM NaCl, 2% DMF
リザーバー	12%-26% PEGmme 3350 0.14 M-0.24 M Sodium citrate 0.10 M Bis-Trispropane pH 6.5
混合比	リザーバー : タンパク質 = 1 $\mu$ L : 1 $\mu$ L
結晶化の温度	20 °C

**Table 4-5-3.** ハンギングドロップ蒸気拡散法による条件検討 2

結晶化方法	ハンギングドロップ蒸気拡散法
タンパク質溶液	3.2 mg/mL PHBH-MWNL, 0.50 mM FAD, 10 mM Tris-HCl (pH 7.4, 4 °C). 100 mM NaCl
リザーバー	14%-17.5% PEGmme 3350 0.14 M-0.24 M Sodium citrate 0.10 M Bis-Trispropane pH 6.5 80 mM 4-hydroxy-3-methylbenzoic acid ( <b>6</b> )
混合比	リザーバー : タンパク質 = 1 $\mu$ L : 1 $\mu$ L
結晶化の温度	20 °C

**Table 4-5-4. PF BL17A による X 線回折強度の測定条件**

BeamLine	PF BL17A
Wavelength (Å)	0.980000
Detector	PILATUS3 S6M
Exposure Time (sec.)	0.2
Oscillation Angle (°)	0.2
Data collection range (°)	0-180°
Detector distance (mm)	392.310
Temperature (K)	100

**Table 4-5-5. PHBH-MWNL-6 複合体の結晶学的データと構造精密化の統計値**

Statistics	PHBH-MWNL in complex with 6
<b>Data collection</b>	
Wavelength (Å)	0.98000
Space group	<i>C</i> 222 <sub>1</sub>
Unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	72.2, 141.9, 89.5
Resolution (Å)	44.7–2.0 (2.07–2.0)
Observed reflections	209,417 (21,416)
Unique reflections	31,316 (3,058)
<i>R</i> <sub>merge</sub> <sup>a</sup>	0.057 (0.867)
Completeness (%)	99.49 (99.29)
<i>I</i> /σ ( <i>I</i> )	19.00 (2.64)
Redundancy	6.7 (7.0)
Molecules/asymmetric unit	1
Data collection facility	BL-17A Photon Factory beamline, Tsukuba, JAPAN
<b>Refinement</b>	
Resolution (Å)	44.7–2.0
Refined reflections	31,289
Free reflections	1,995
<i>R</i> <sub>work</sub> <sup>b</sup> / <i>R</i> <sub>free</sub> (%)	21.77 / 25.35
Ramachandran Plot	
Favored (%)	97
Allowed (%)	0
Outliers (%)	0
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.71
Protein Data Bank code	8Y2S

<sup>a</sup>  $R_{\text{merge}} = \sum |I_{\text{avg}} - I_i| / \sum I_i$ , where  $I_i$  is the observed intensity and  $I_{\text{avg}}$  is the average intensity of observations of symmetry-related reflections.

<sup>b</sup>  $R_{\text{factor}} = \sum |F_p - F_{p(\text{calc.})}| / \sum F_p$ , where  $F_p$  and  $F_{p(\text{calc.})}$  are observed and calculated structure factors;  $R_{\text{work}}$  and  $R_{\text{free}}$  are calculated with refined and free reflections, respectively.

## 第六節 引用文献

- (1) Cole, L. J.; Gatti, D. L.; Entsch, B.; Ballou, D. P. Removal of a Methyl Group Causes Global Changes in *p*-Hydroxybenzoate Hydroxylase. *Biochemistry* **2005**, *44*, 8047-8058.
- (2) Eppink, M. H.; Schreuder, H. A.; Van-Berkel, W. J. Structure and function of mutant Arg44Lys of 4-hydroxybenzoate hydroxylase Implications for NADPH binding. *Eur. J. Biochem.* **1995**, *231*, 157-165.
- (3) Wang, J.; Ortiz-Maldonado, M.; Entsch, B.; Massey, V.; Ballou, D.; Gatti, D. L. Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 608-613.
- (4) Lah, M. S.; Palfey, B. A.; Schreuder, H. A.; Ludwig, M. L. Crystal Structures of Mutant *Pseudomonas aeruginosa* *p*-Hydroxybenzoate Hydroxylases: The Tyr201Phe, Tyr385Phe, and Asn300Asp Variants. *Biochemistry* **1994**, *33*, 1555-1564.
- (5) Gatti, D. L.; Palfey, B. A.; Lah, M. S.; Entsch, B.; Massey, V.; Ballou, D. P.; Ludwig, M. L. The mobile flavin of 4-OH benzoate hydroxylase. *Science* **1994**, *266*, 110-114.
- (6) Gatti, D. L.; Entsch, B.; Ballou, D. P.; Ludwig, M. L. pH-Dependent Structural Changes in the Active Site of *p*-Hydroxybenzoate Hydroxylase Point to the Importance of Proton and Water Movements during Catalysis. *Biochemistry* **1996**, *35*, 567-578.
- (7) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.
- (8) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204–1214.
- (9) Schreuder, H.A.; van der Laan, J. M.; Hol, W. G.; Drenth, J. Crystal Structure of *p*-Hydroxybenzoate Hydroxylase Complexed with its Reaction Product 3,4-Dihydroxybenzoate. *J. Mol. Biol.* **1988**, *199*, 637-648.



(10) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

(11) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501.

(12) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 352–367.

## 第五章 総括

### 第一節 本研究の総括

本研究では、機械学習を用いた **FMO** の効率的な機能改変に成功した。**FMO** は天然において多種多様な天然物の生合成に関与するため、非常に重要な酵素ファミリーであり、**FMO** の機能改変による天然物誘導体の効率的な創出が期待される。しかし、**FMO** は酵素反応時のコンフォメーション変化に富むため、結晶構造をもとにしても酵素反応時の活性部位の形を予想することが困難であり、効率的な機能改変が難しかった。さらに、酵素の配列には膨大なパターンが存在し、実験室レベルでは検証できる変異体酵素の数は非常に限られている。本研究では、**Val47, Trp185, Leu199, Leu210** の 4 アミノ酸残基について部位特異的な変異を導入した。そして、 $20^4=160,000$  種類の変異体の中の僅か **165** 種類 (**0.1%**程度) を教師データとして使用し、機械学習をすることで、高活性変異体が濃縮されたセカンドライブラリーを創出することに成功した。また、セカンドライブラリーで最も活性が高かった **PHBH-MWNL** は **6** に対して、野生型酵素の **100** 倍以上の活性を持っていることが判明した。

**PHBH-MWNL** について、活性が上昇した詳細なメカニズムを解明するために、結晶化条件の探索をすることで、単結晶を獲得することに成功した。結晶構造解析をすることにより、**PHBH-MWNL-FAD-6** 複合体の結晶構造を獲得することに成功した。獲得された結晶構造を詳細に解析することにより、**L199** が **Asn** 残基に置換されたことにより、水素結合を形成することで、水分子を安定に留めておくことができるようになり、プロトンリレーの安定化に繋がったことが推測される。通常酵素の機能改変では、基質結合部位の形状にだけ注目しがちであるが、本研究から、プロトンリレーの安定化による酵素の機能改変の有効性が示唆されたため、酵素の機能改変の幅が広がることが期待される。

さらに、機械学習により予測されたセカンドライブラリーは、初期のライブラリーに比べて、アンカップリングが少ない高活性変異体が濃縮されていることが判明した。近年、合成生物学が注目されており、**FMO** を宿主となる微生物に発現させることで、有用な化合物の効率的な生産が期待される。しかし、**FMO** のアンカップリングは微生物の生命維持に必要な **NADPH** を浪費し、さらに、アンカップリングにより放出される過酸化水素は微生物にとって有毒であり、宿主となる微生物の生育の妨げになることが懸念される。このことから、機械学習によるアンカップリングの少ない **FMO** の創出は、**FMO** の産業的な応用を促進し、効率的な天然物誘導体の創出が可能になることが期待される。

## 第二節 今後の展望

本研究で、機械学習が **FMO** の機能改変に有効であることが示され、本手法を他の研究に応用することにより、天然物誘導体の効率的な取得が可能になることが期待された。しかし、課題としては、教師データの収集にかなりの労力が要することが挙げられる。また、上述のように **FMO** に独自のアンカップリングにより、**NADPH** の消費量をもとにしたハイスループットなアッセイが困難である。さらには、コンフォメーション変化に富むため、*apo* 体と *holo* 体で基質結合部位の構造が変わるため、ドッキングシミュレーションを用いても正確な酵素と基質の親和性を計算することが困難である。そこで、解決策として「**MD** シミュレーション」により教師データを収集し、「弱機械学習」を使用することを現在検討している。

従来のドッキングシミュレーションは、酵素の構造を剛体として扱い、シミュレーションをするので、**FMO** のようなフレキシブルな酵素には不向きである。それに対して、**MD** シミュレーションでは、酵素のコンフォメーション変化を考慮したシミュレーションをすることができるため、**FMO** とより正確に相互作用の強さを定量できる可能性がある。例えば、**FMO** の一種である **TropB** はベンゼン環を脱芳香化する酵素活性を持ち、本来の基質だけでなく、様々な基質誘導体との活性試験が行われてきた。そして、**TropB** が高い活性を示した基質誘導体との相互作用について、**MD** シミュレーションを用いることで、その基質認識機構について詳細な考察がされていた<sup>1)</sup>。

また、機械学習は教師データの精度に大きく影響されるが、精度が高いデータを集めることが難しい状況も存在する。そのような状況には、弱機械学習が有効であることがある。弱機械学習では、精度が弱めの情報が多く手に入るときに有効な手法であり、**MD** シミュレーションで集めた教師データを用いて、弱機械学習が有効になる可能性がある。例えば、**MD** シミュレーションによるタンパク質の安定性予想をし、それを用いて野生型よりも安定なタンパク質の変異体を創出した例が存在する<sup>2)</sup>。機械学習はタンパク質の機能改変に有効であることが示唆され、今後も機械学習を用いた酵素の機能改変が益々行われていくことが期待される。

### 第三節 引用文献

(1) Benítez, A. R.; Tweedy, S.; Baker-Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L. 3rd.; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H. Structural basis for selectivity in flavin-dependent monooxygenase-catalyzed oxidative dearomatization. *ACS Catal.* **2019**, *9*, 3633–3640.

(2) Jokinen, E.; Heinonen, M.; Lähdesmäki, H. mGPFusion: predicting protein stability changes with Gaussian process kernel learning and data fusion. *Bioinformatics* **2018**, *34*, i274–i283.

## 謝辞

研究を行うにあたり、幅広い面でご指導頂きました、渡辺賢二教授、岸本真治講師に厚く御礼申し上げます。

機械学習について全面的に指導して下さった、産総研の齋藤裕主任研究員に厚く御礼申し上げます。

タンパク質の X 線結晶構造解析において全面的にご指導頂きました、生命物理化学研究室の原幸大准教授に厚く御礼申し上げます。

X 線回折実験をして下さった SLS 及び PF のスタッフの方々に厚く御礼申し上げます。