Mining and engineering natural-product biosynthetic pathways

LITERATURE SEMINAR#1 18-12-25

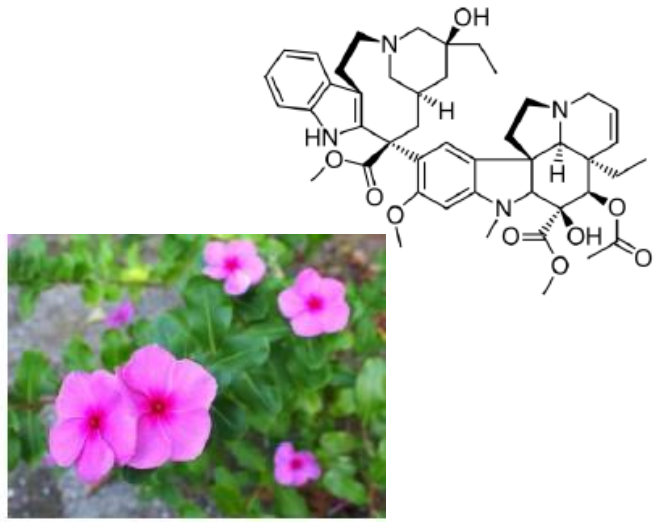
触媒医療G Y.KAMIMURA

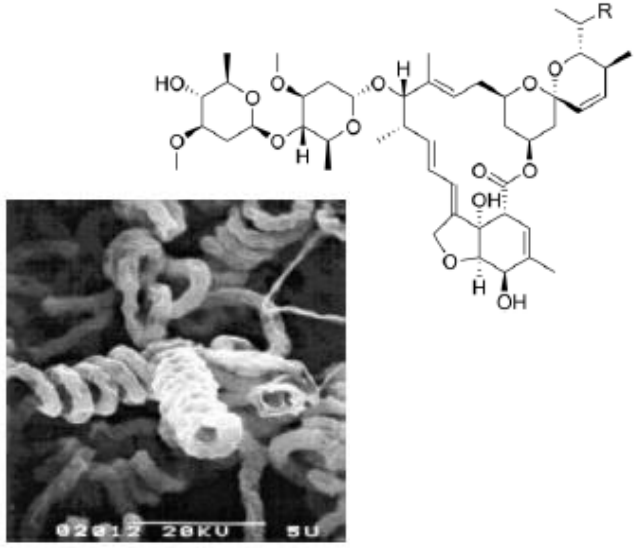
1. Introduction
    - Secondary metabolites
2. Current progress of the field
    - Genome mining approach
    - Enzyme engineering approach
3. Summary

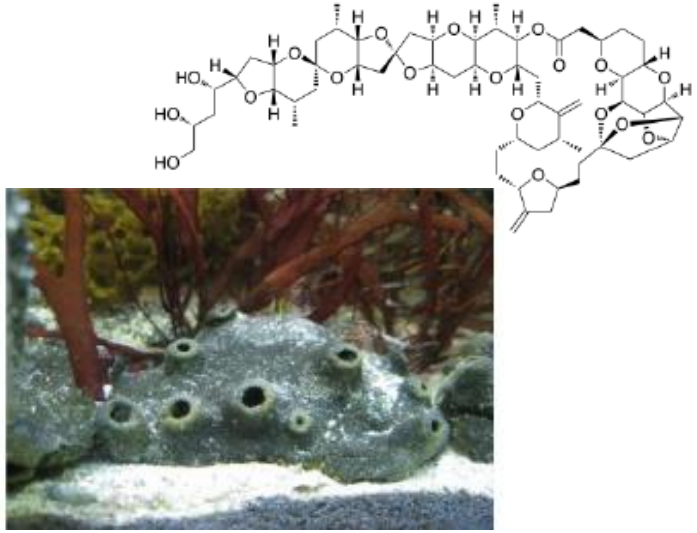
Appendix

Secondary metabolites

- Compounds created by bacteria, fungi, plants, etc.
- Usually, they are not directly involved in normal growth, development, or reproduction. (Vaishnav and Demain 2011)
- They have multifarious structure, and bioactivity.



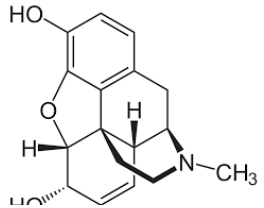


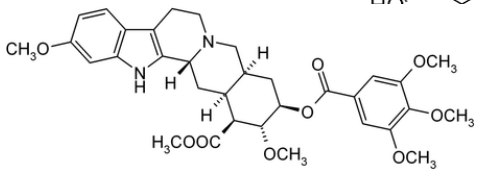


Structural diversity and classification

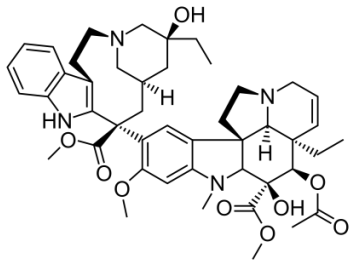
Alkaloids

Morphine



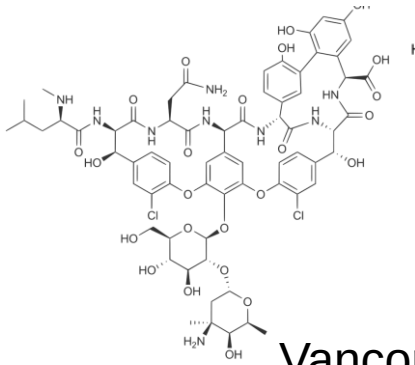


Reserpine

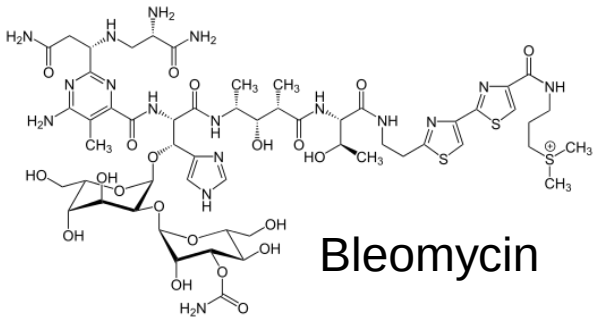


Vinblastine

Nonribosomal peptides

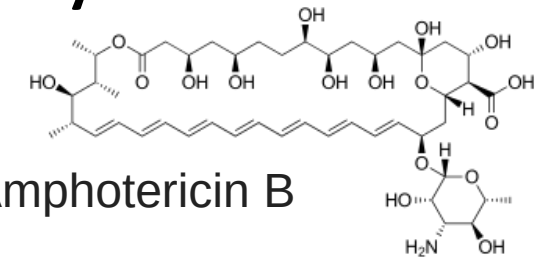


Vancomycin

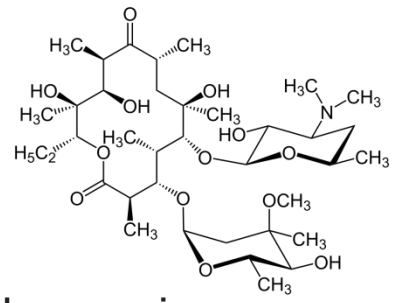


Bleomycin

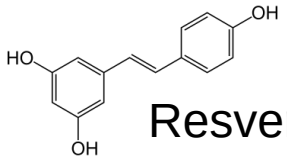
Polyketides



Amphotericin B

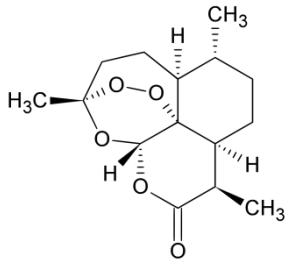


Erythromycin



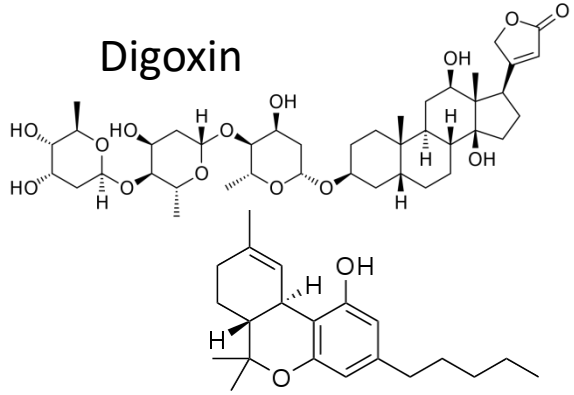
Resveratrol

Terpenoids



Artemisinin

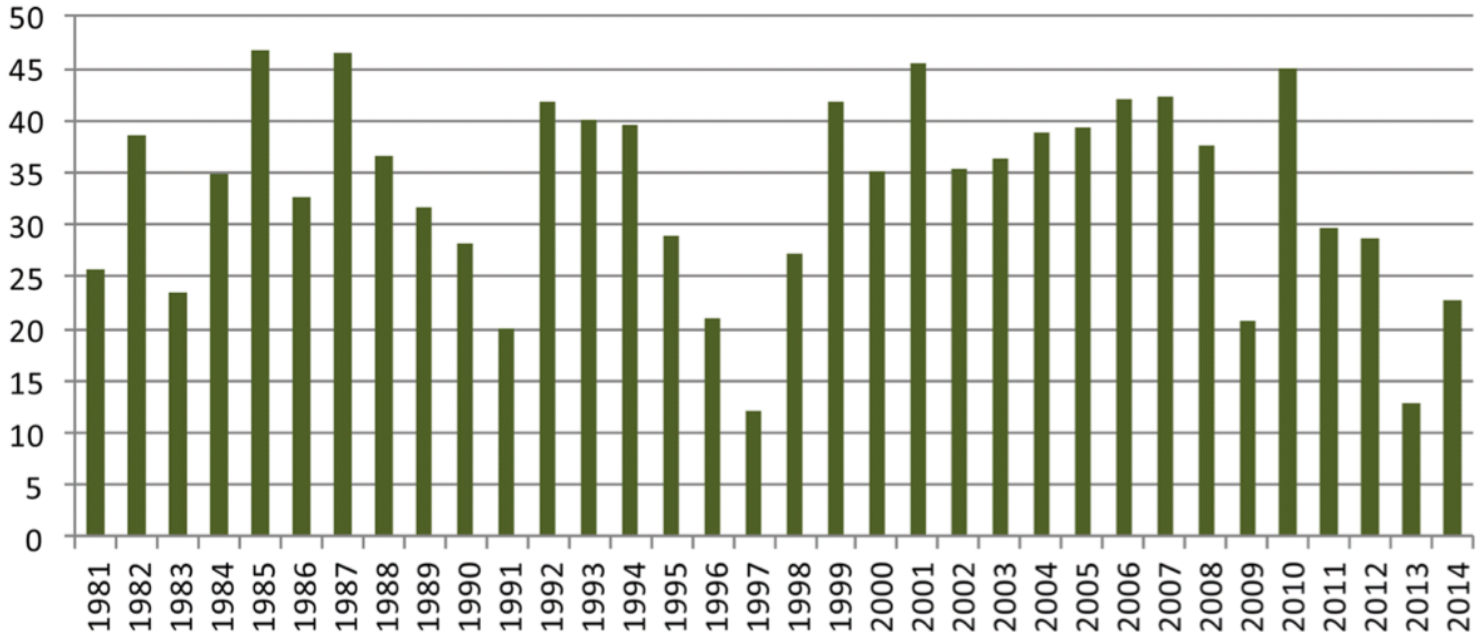
Digoxin



Tetrahydrocannabinol

Natural products as a drug

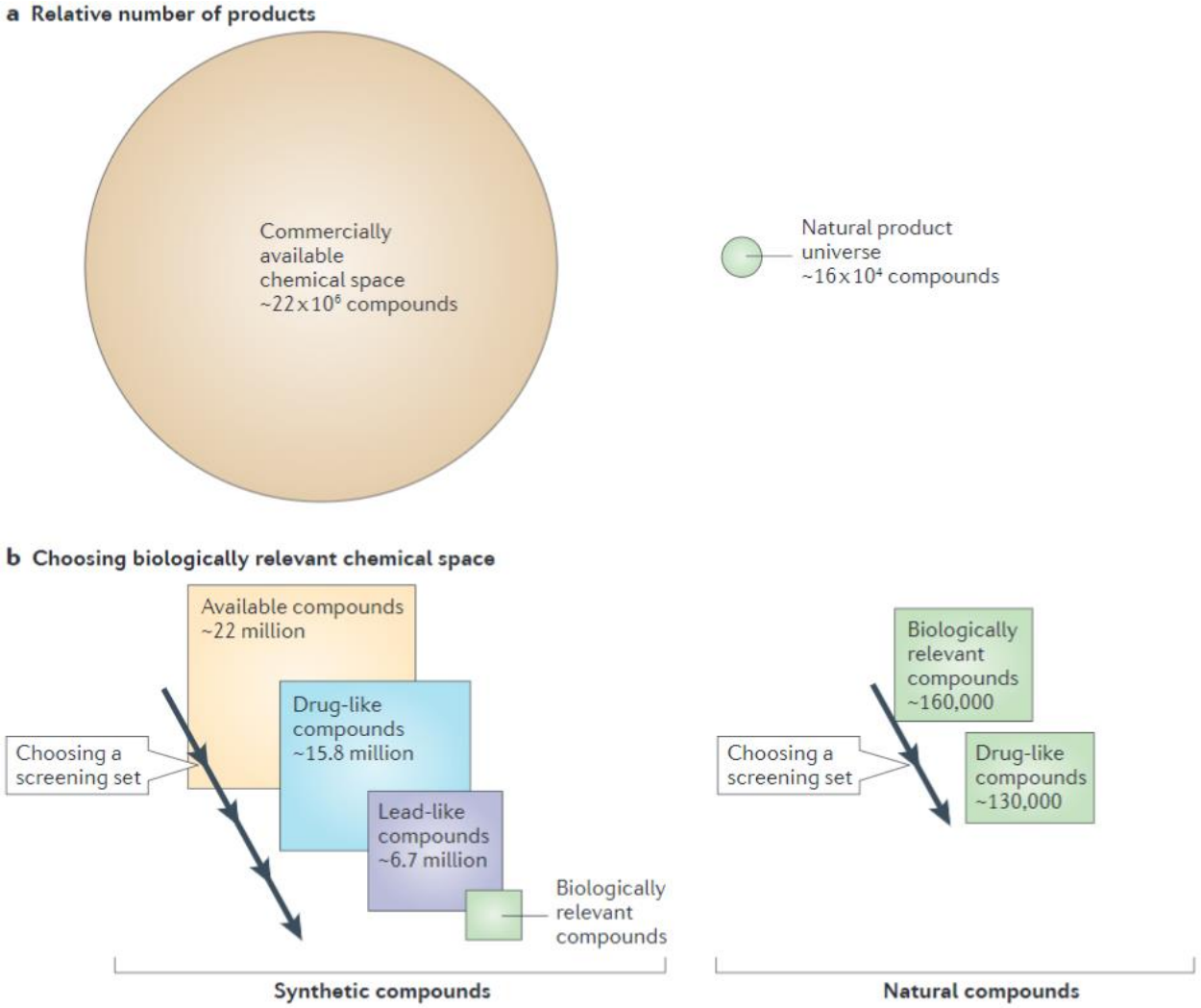
- NPs have been continuously exploited for drug discovery.



Percentage of NP origin drug in newly approved drug
(From : Newman and Cragg , *J. Nat. Prod.* **2016**, *79*, 629−661)

Biologically relevant chemical space

- In drug discovery, biologically relevant chemical space is more important than just a library size.
- NPs have wide range of pharmacophores, and high degree of stereochemistry.
- The chemodiversity of NPs is much larger than those of synthetic compounds. (Feher and Schmidt, 2003)

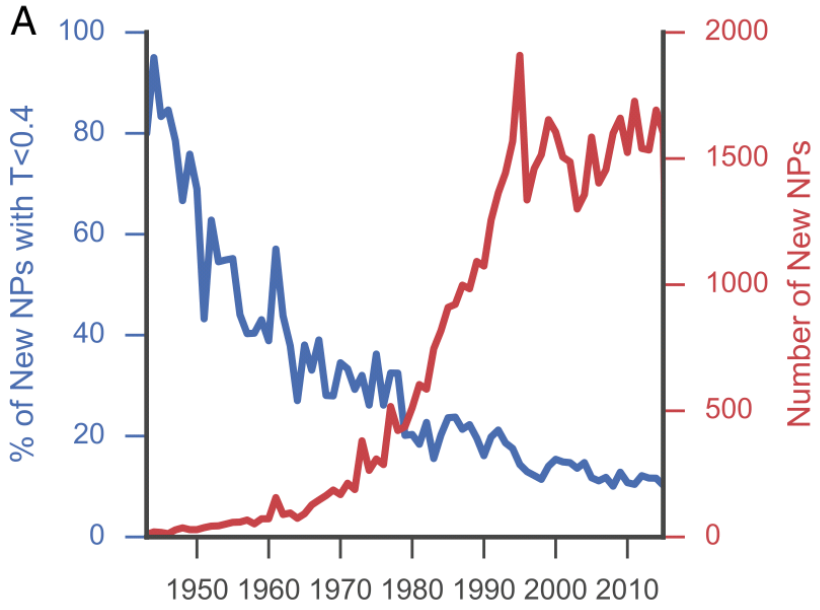


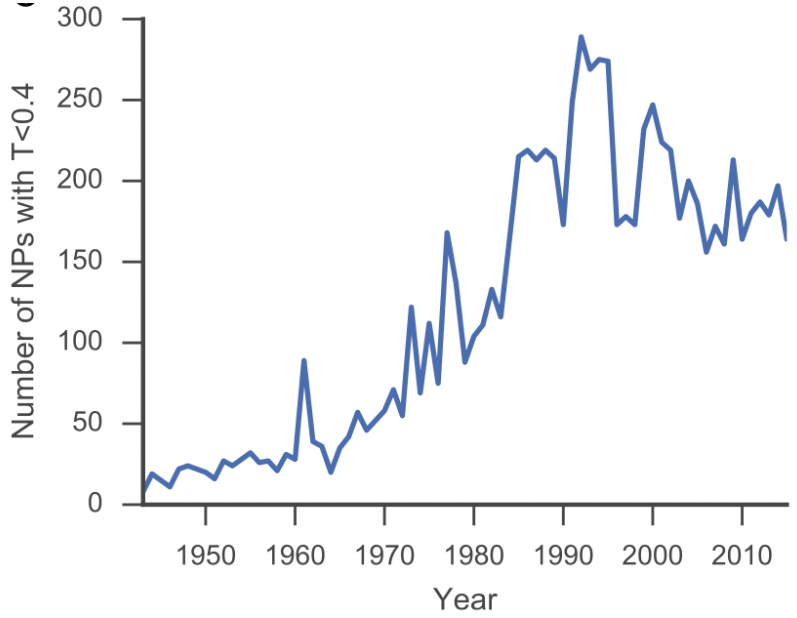
Hervey et al., Nat. *Rev. Drug Discov.*, **2015**, *14*, 111-129
Feher and Schmidt, *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 218-227

Number and diversity of new NPs

However, number and diversity of new NPs seem to reached a plateau.

➡Method to discover structurally new NPs, or artificially diversify NPs are needed.





1. Introduction
    - Secondary metabolites
2. Current progress of the field
    - Genome mining approach
    - Enzyme engineering approach
3. Summary

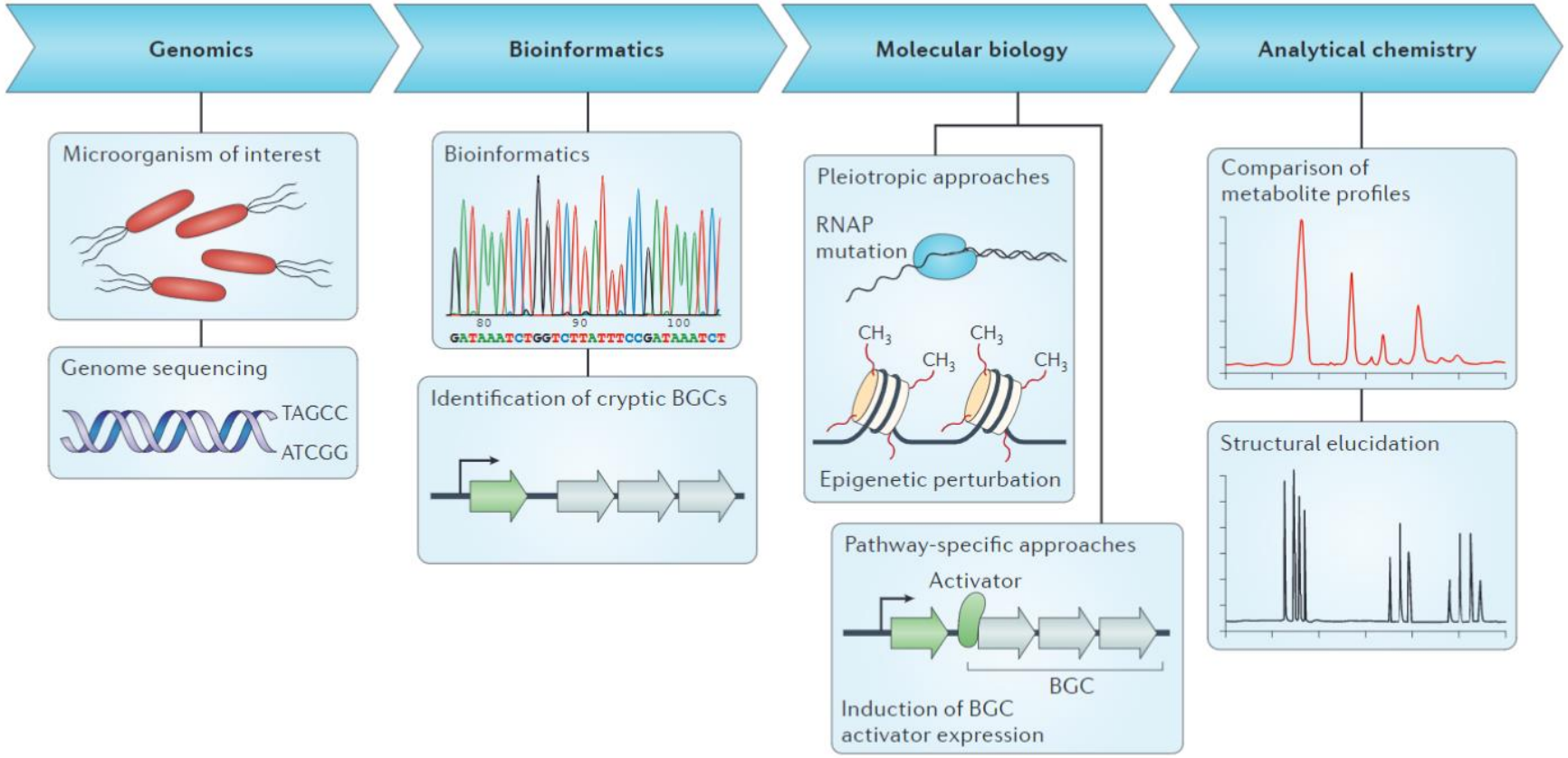
Appendix

Genome mining approach

- Over the past decade many biosynthetic gene clusters (BGCs) have been uncovered.
- However, most gene clusters are silent under laboratory conditions, presenting a bottleneck for natural product discovery.
- Bio-informatics-based analysis predicted that several cryptic BGCs in many species are likely to encode products with novel structures.

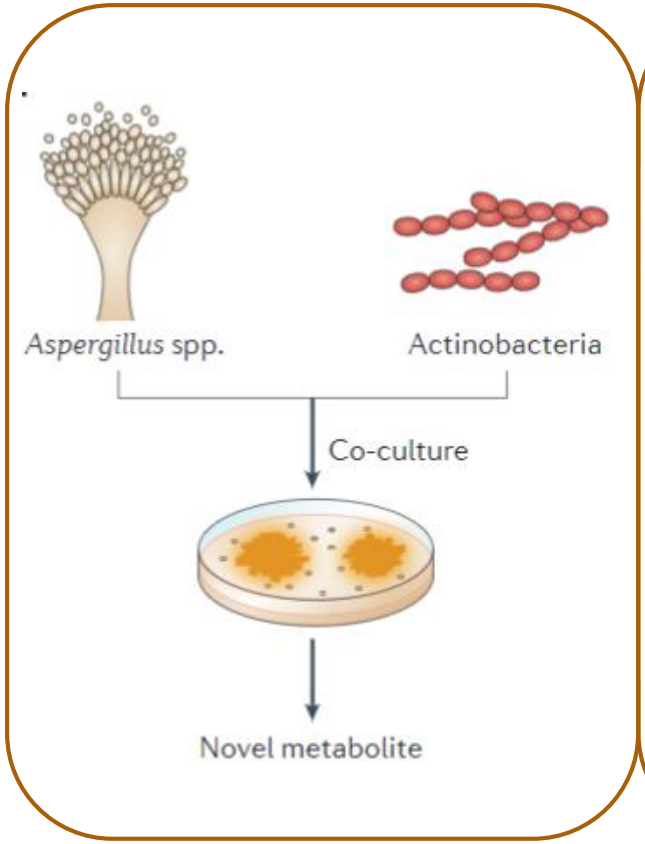
Genome mining approach

- Thus, the activation of silent gene clusters will become a new direction of natural product discovery.

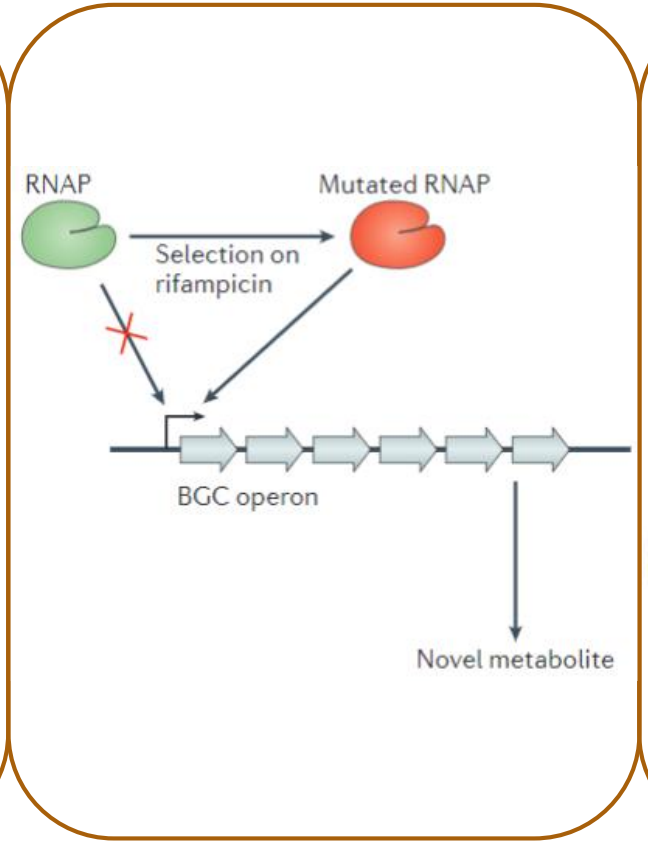


Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.

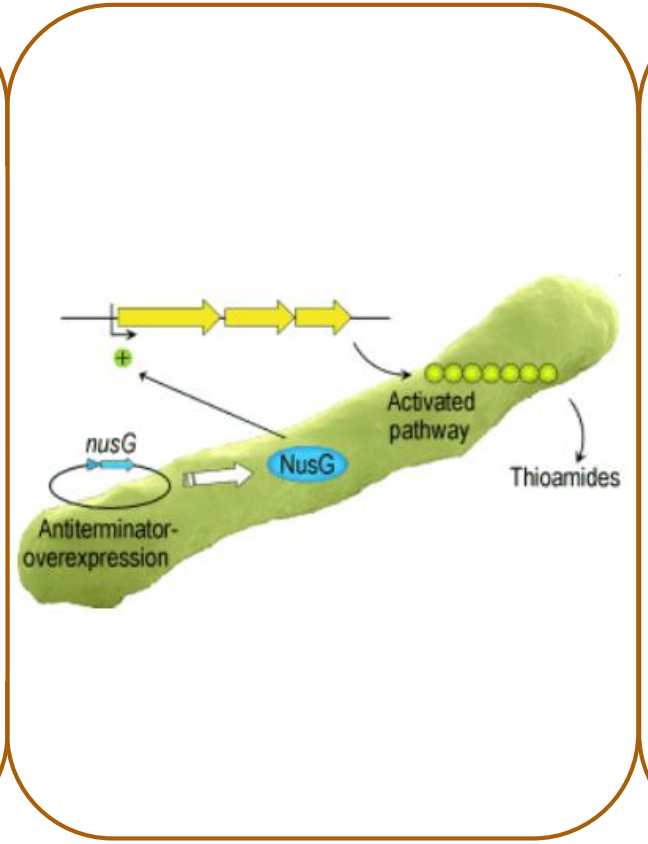
Pleiotropic methods



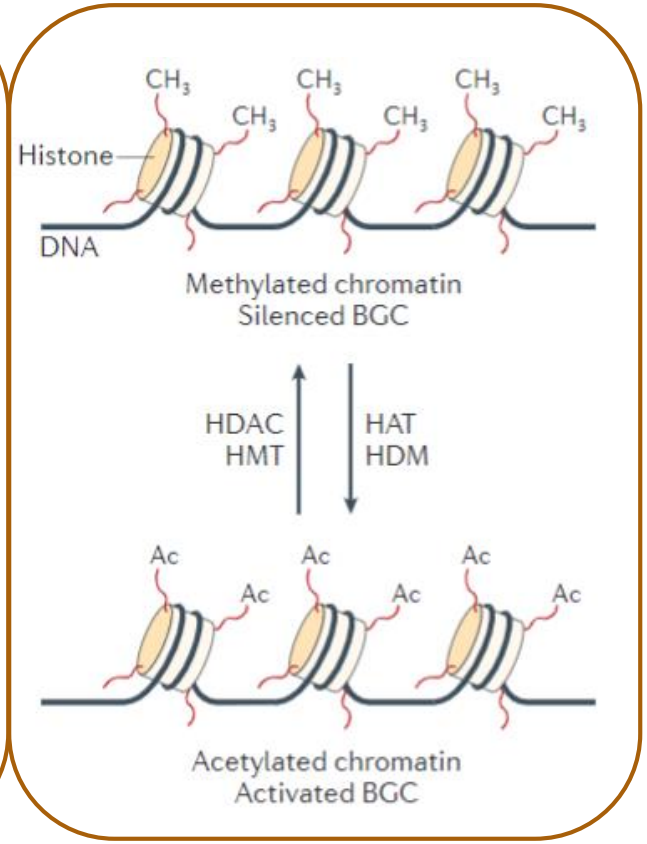
Variation in growth condition



Engineering the transcription and translation machinery



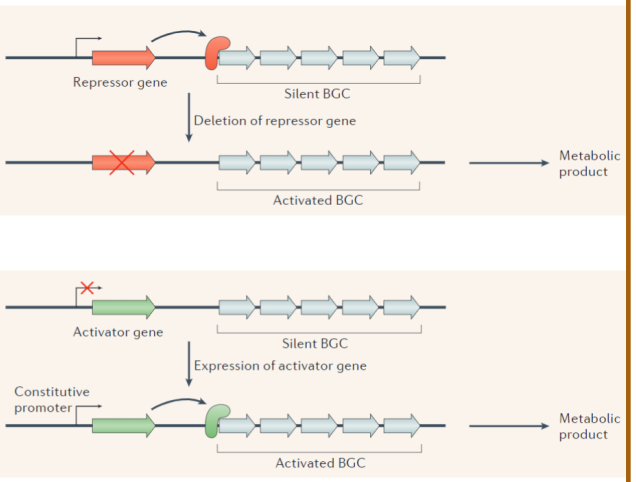
Manipulating global regulators



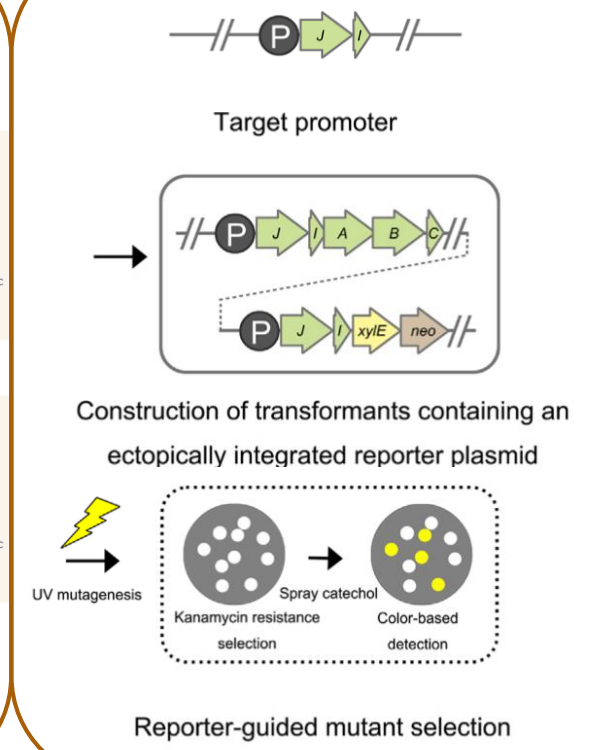
Epigenetic perturbation

Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.

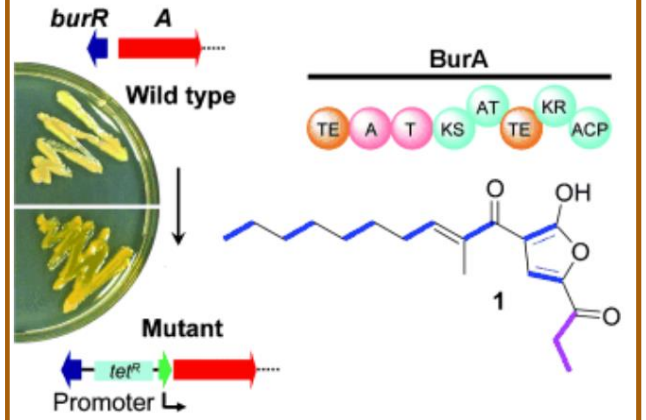
Pathway-specific methods



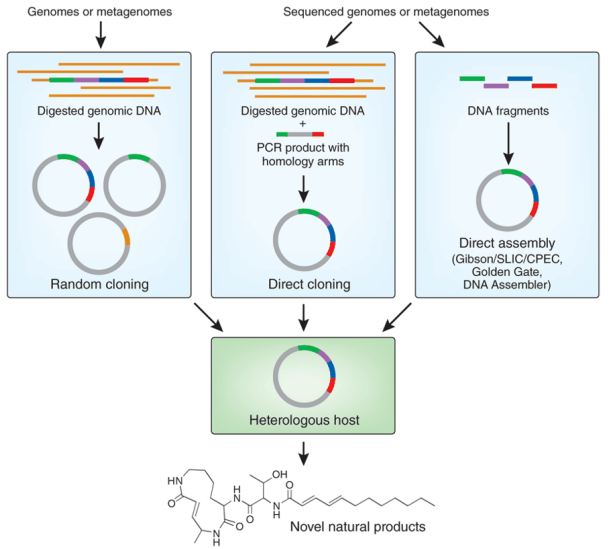
Manipulating pathway specific regulators



Reporter-guided mutant selection



Refactoring



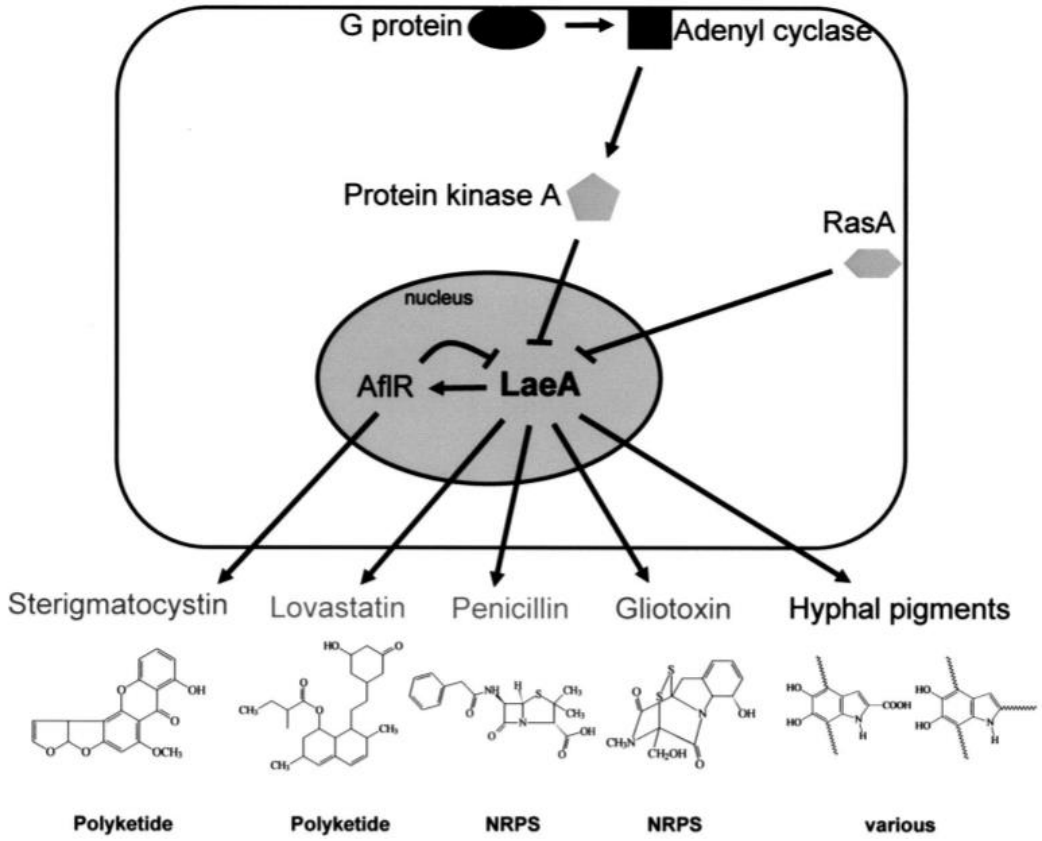
Heterologous expression

Example of epigenetic approach

Pioneer work in epigenetic perturbation approach

- Bok and Keller found LeaA, which activate the expression of transcription factor *aflR.*
- By detailed analysis, LeaA was found to be
  - ・Global regulation factor conserved in filamentous fungi.
  - ・Concerned to regulation of gene expression.
  - ・Bearing binding site of *S*-adenosyl methionine.
  - ・Localizes in nuclei.

➡LeaA = regulating expression by PTM??



Bok and Keller et al., *Eukaryotic Cell*, **2004**, *3*, 527-535.
Asai and Oshima, *生化学*, **2016**, *88*, 643-648.

Pioneer work in epigenetic perturbation approach

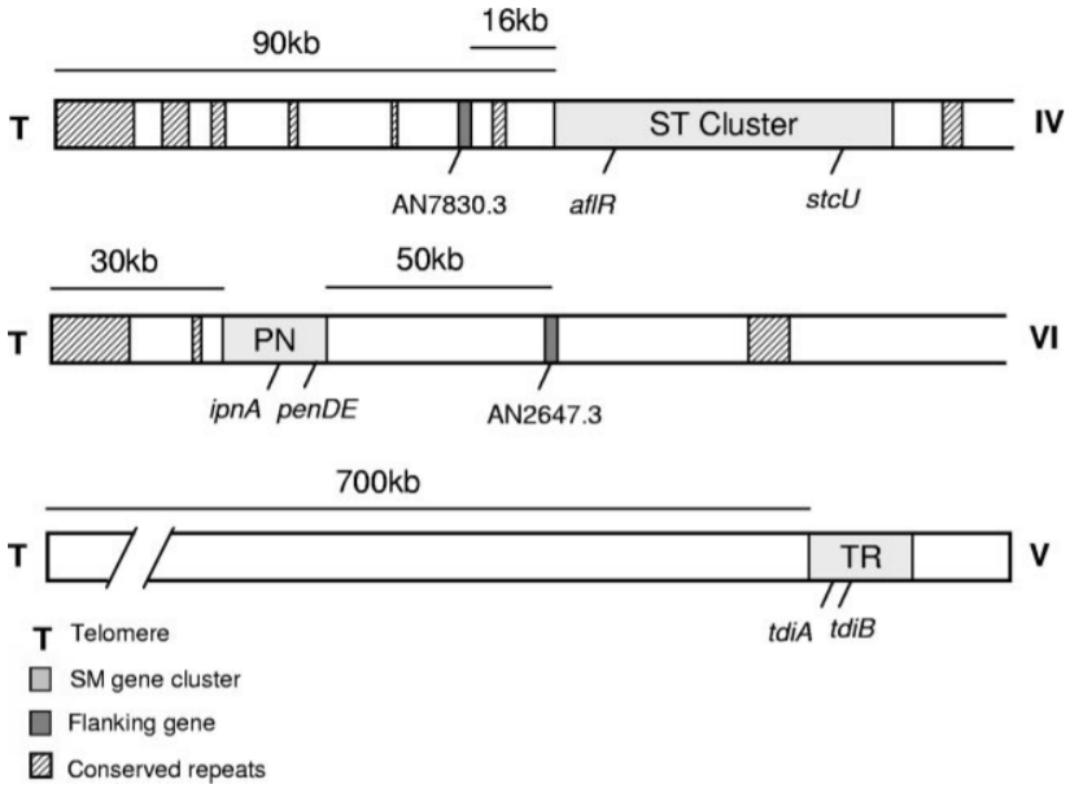
- Although farther research elucidated that LeaA does not function as methylation enzyme, Bok's work connected the secondary metabolism and epigenetic regulation.

- Another basis : Many of the BGCs of filamentous fungi are located near the telomere (Nierman et al., 2005)(Rehmeyer et al., 2006).
- To confirm the relation, Keller's group conducted further research.

Shwab et al., *Eukaryotic Cell*, **2007**, *6*, 1656-1664
Asai and Oshima, *生化学*, **2016**, *88*, 643-648.

Pioneer work in epigenetic perturbation approach

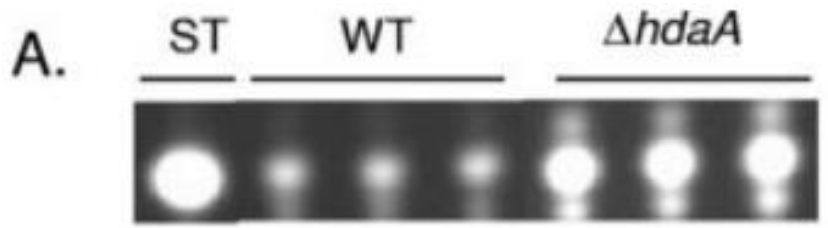
- Shwab et al. (2007) examined the effect of HDAC loss on the 3 best characterized BGCs in *Aspergillus nidulans.*
  - ST : sterigmatocytin cluster
  - PN : penicillin cluster
  - TR: terrequinone A cluster
- ST, PN is in subtelomeric region, but TR is not.



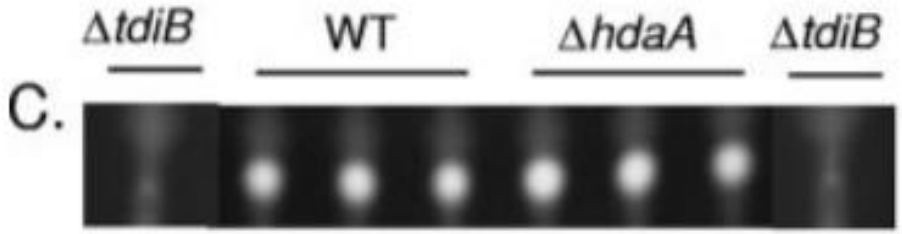
Pioneer work in epigenetic perturbation approach

- They created isogenic lines losing one or more of *hdaA, hosB, hstA.* (HDACs)
- In *ΔhdaA,* production of ST and PN increased, but not TR.

ST



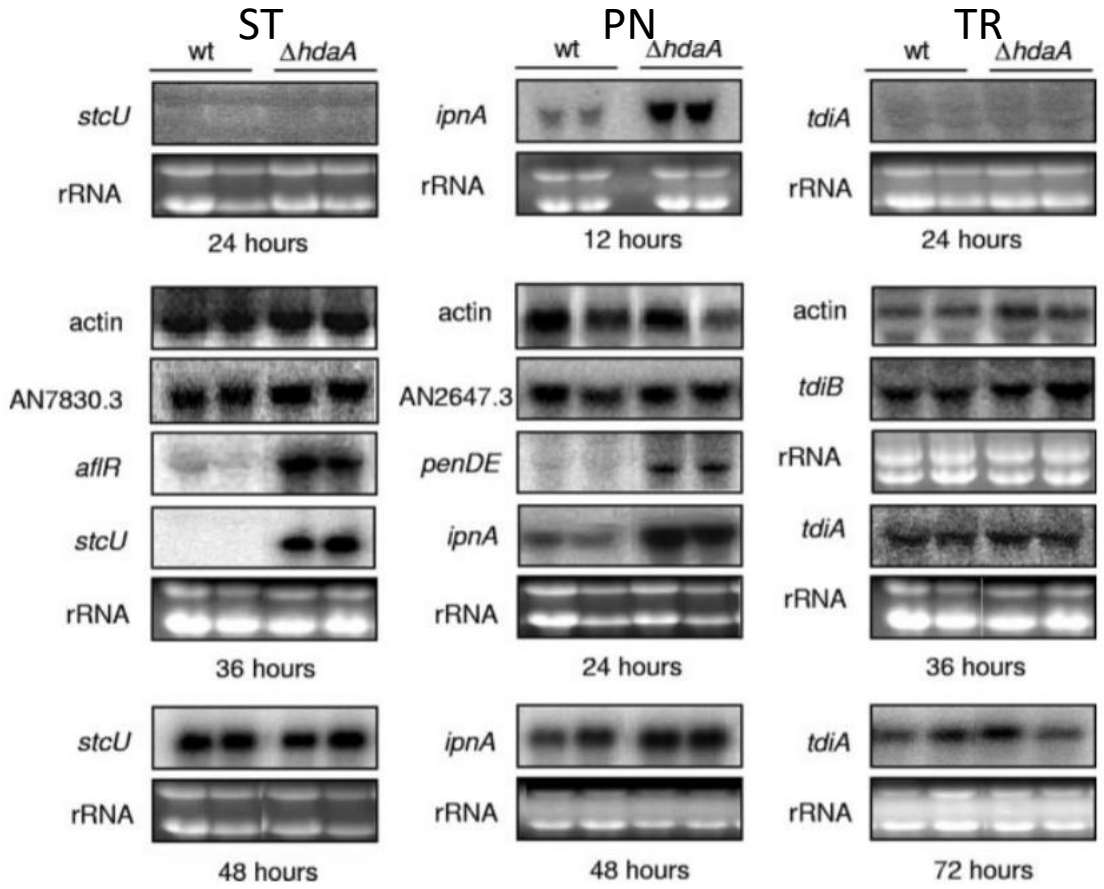
TR

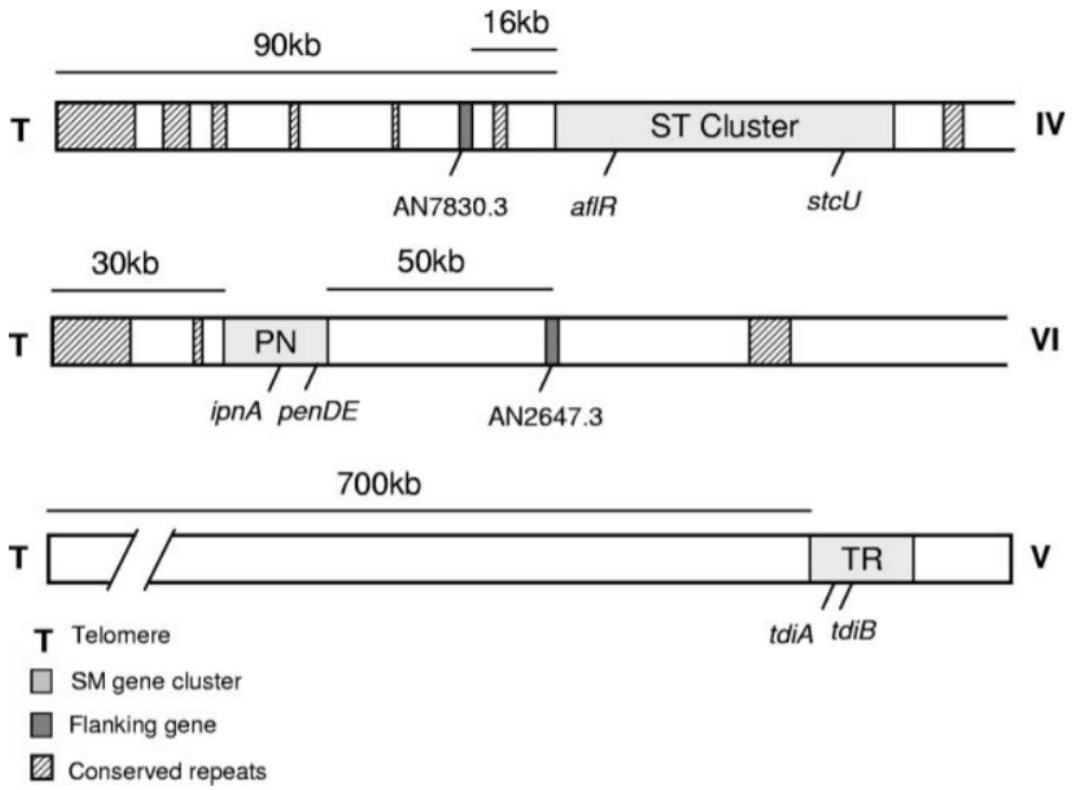


Shwab et al., *Eukaryotic Cell*, **2007**, *6*, 1656-1664

Pioneer work in epigenetic perturbation approach

- They picked representative genes to see the change in mRNA production.

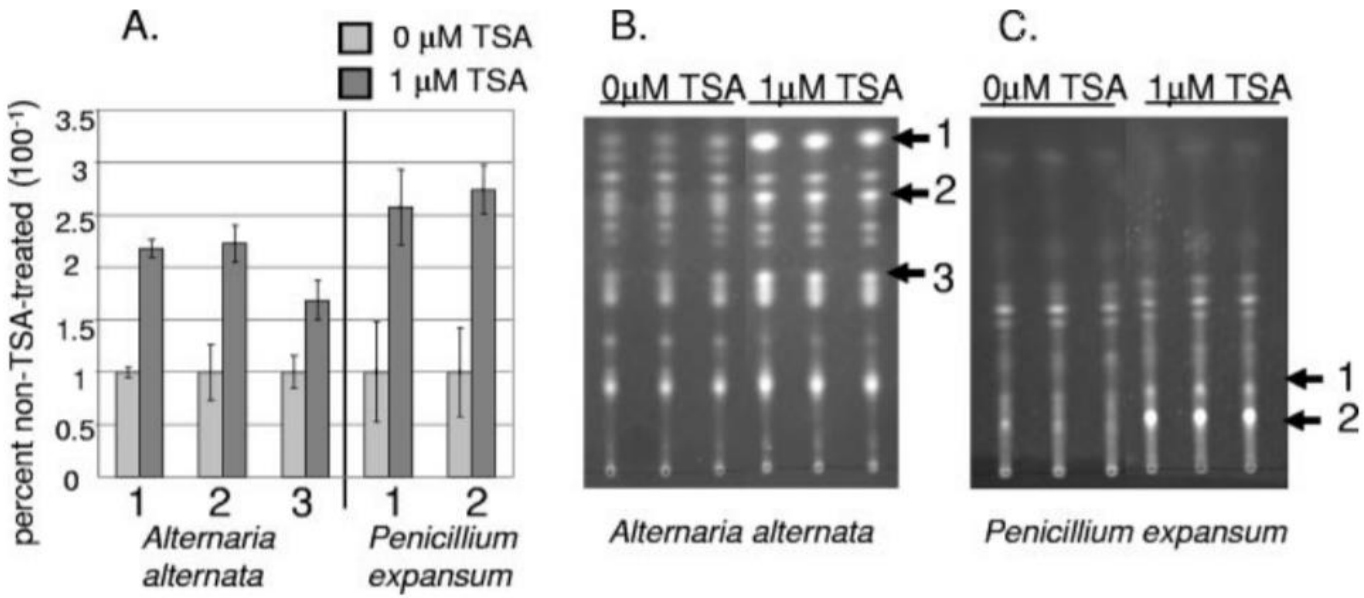




Shwab et al., *Eukaryotic Cell*, **2007**, *6*, 1656-1664

Pioneer work in epigenetic perturbation approach

- To see the generality of HDAC regulation, A.alternate and P.expansum was treated with HDAC inhibitor TSA.



Genetic engineering / **Chemical** approach are possible.

Shwab et al., *Eukaryotic Cell*, **2007**, *6*, 1656-1664

Genetic engineering approach

Bok et al., 2009

Induced loss-of-function CclA (involved in H3K methylation) in *A. nidulans*

➡Activation of cryptic BGCs, and generation of several metabolites which was not previously known to be produced by *A. nidulans*

cclA and *hdaA* are highly conserved in filamentous fungi

➡These methods are applicable to various fungi.

e.g. : *Calcarisporium arbuscula*(Mao et al., 2015)

Pestalotiopsis fici(Wu et al., 2016)

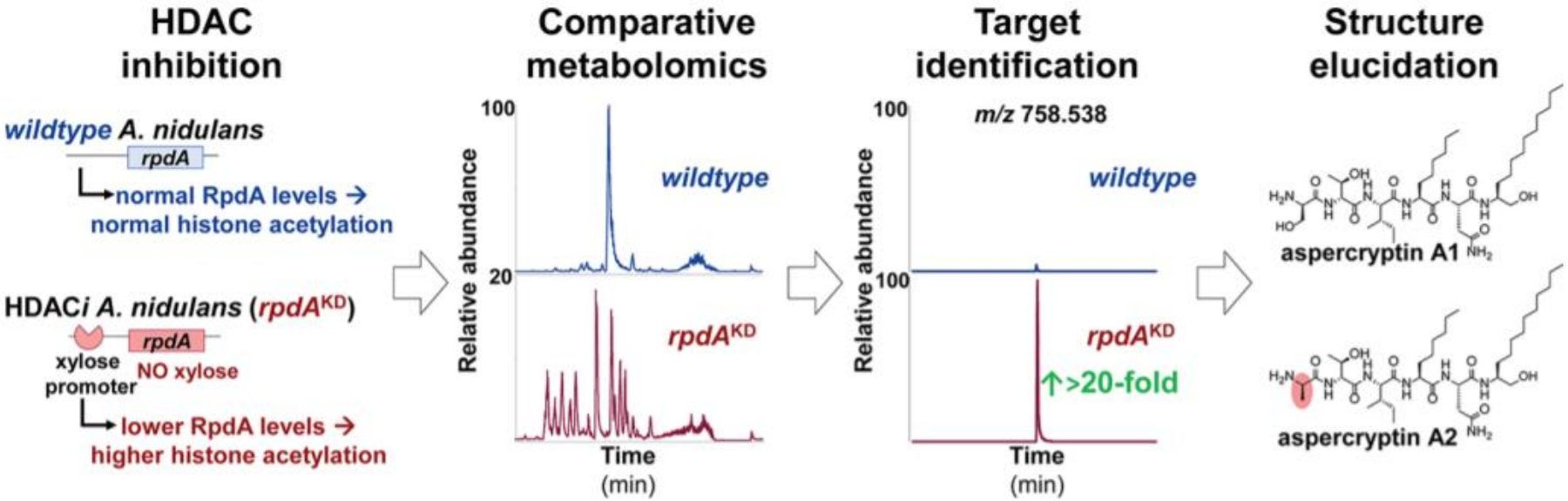
Bok et al., *Nat. Chem. Biol.*, **2009**, *5*, 462-464.
Mao et al., *Angew. Chem. Int. Ed.*, **2015**, *54*, 7592-7596, Wu et al., *Org. Lett.*, **2016**, *18*, 1832-1835

Genetic engineering approach

Henke et al., 2016

・PTM enzymes required for the survival is impossible to delete.

➡Replaced the promoter with xylose promoter

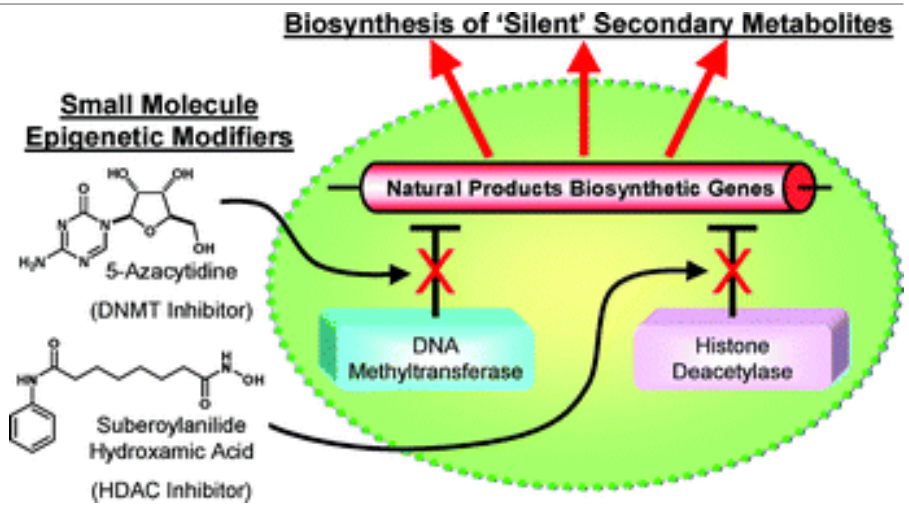


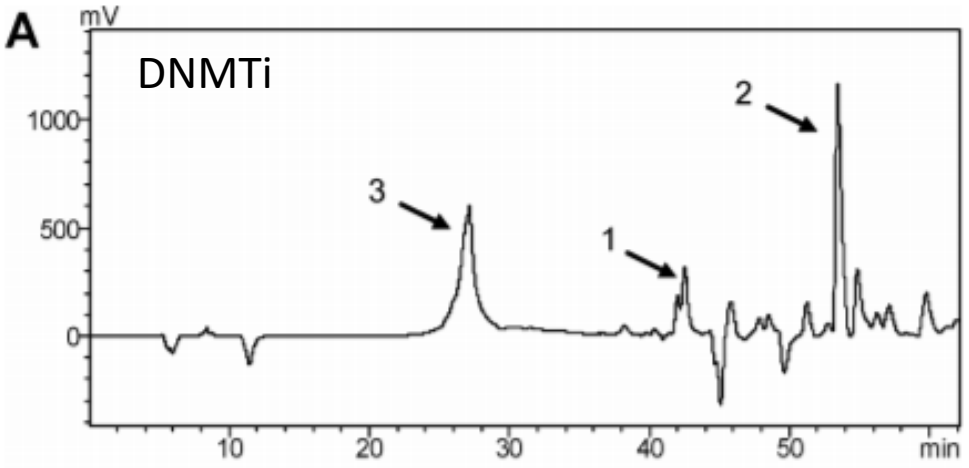
Chemical approach

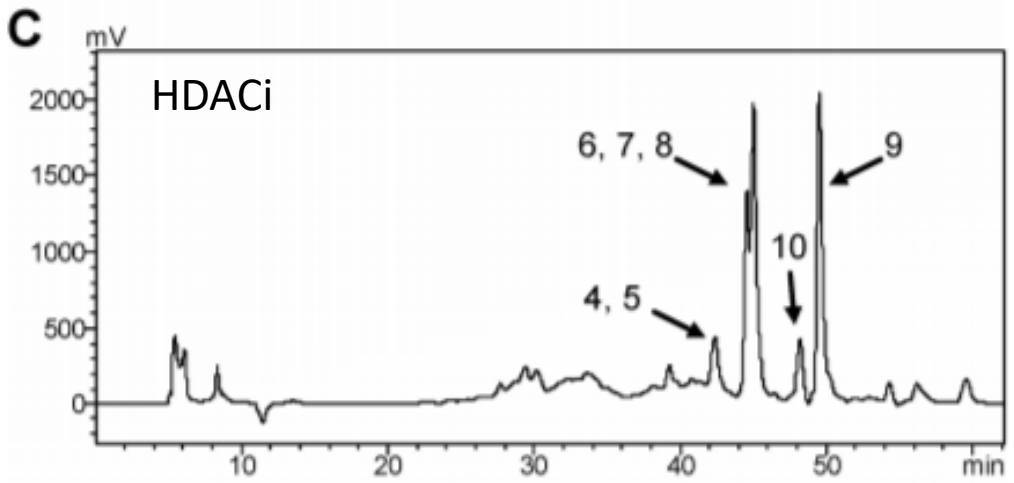
- Williams et al., 2008

Treated Cladosporium Cladosporioides with DMTs or HDAC inhibitor

➡Change in metabolic profile, and generation of new chladochrom.







Chemical approach

Henrikson et al., 2008

Treated *Aspergillus niger* with the same HDAC inhibitor as Williams

➡Discovered novel metabolite with unprecedented structure.

Asai et al., 2015

Combined the chemical approach with semi-synthetic approach to produce novel compound with bioactivity.

Henrikson et al., *Org. Biomol. Chem*.,**2009**, *7*, 435-438
Asai et al., *Nat. Chem*.,**2015**, *7*, 737-743

Limitations of epigenetic perturbation approach

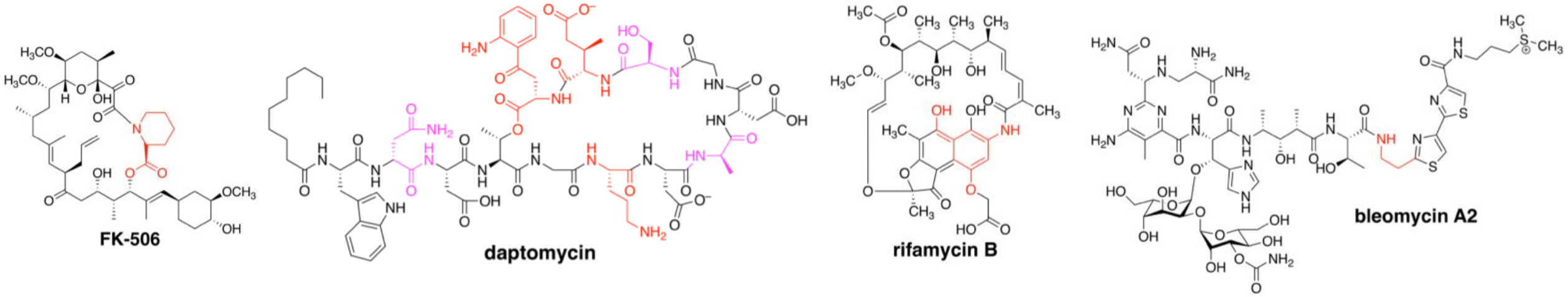
- Application of genetic engineering approach is limited to few strains.
  - Genome information, and establishment of transformation method is necessary.
- Predictive change is difficult using pleiotropic approach, such as HDAC inhibition by drug.
- BGCs that are not strongly regulated by epigenetic regulation are difficult to upregulate by this method.

1. Introduction
    - Secondary metabolites
2. Current progress of the field
    - Genome mining approach
    - Enzyme engineering approach
3. Summary

Appendix

Nonribosomal peptides (NRPs)

- Structure of NRPs are made by amino acids, including nonproteinogenic amino acids.

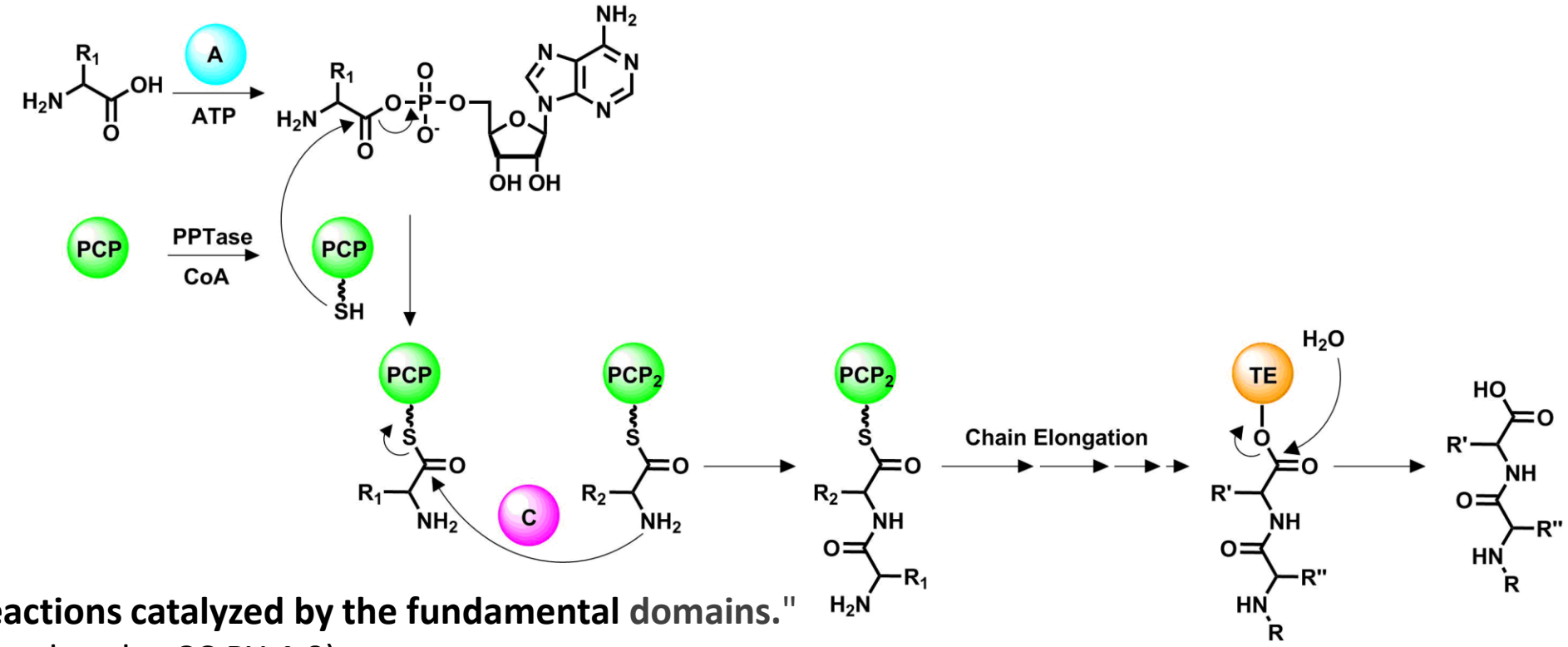


NRPs

(From Kudo et al, *J. Ind. Microbiol. Biotechnol.*, **2018**, not yet assigned to an issue.)

Nonribosomal peptides (NRPs)

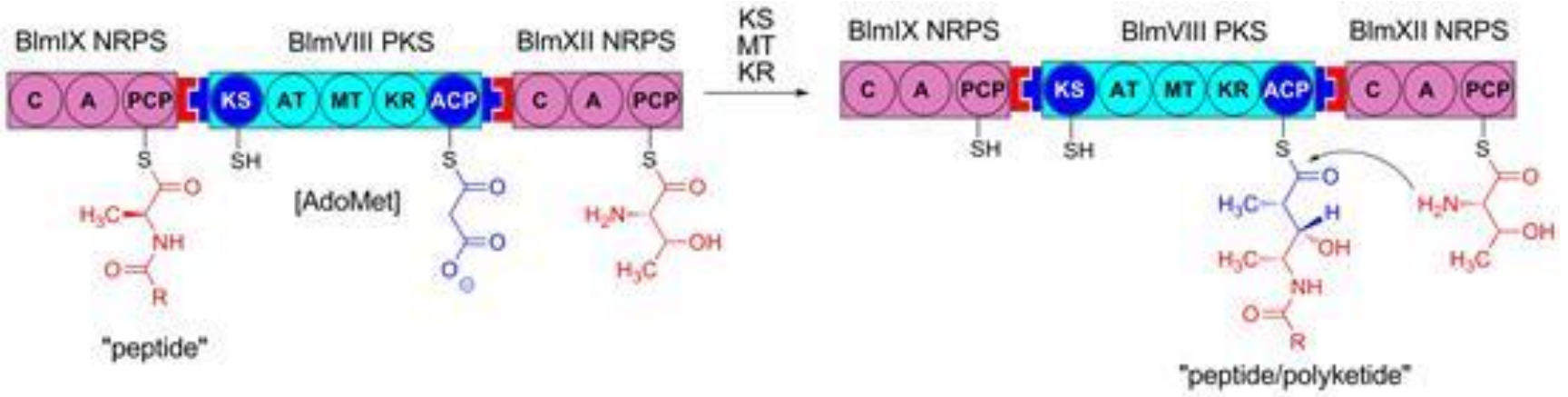
lNRPs are assembled by NRP synthetases, which consist of adenylation domain and peptidyl carrier protein, and other domains for modification.



"The NRPS reactions catalyzed by the fundamental domains."

Nonribosomal peptides (NRPs)

- NRPs get various modification including reduction, oxidation, cyclization, epimerization, methylation, etc.
- Enzymes are modularized.
- Sometimes NRPS and PKS are combined to form NRP-PKs.

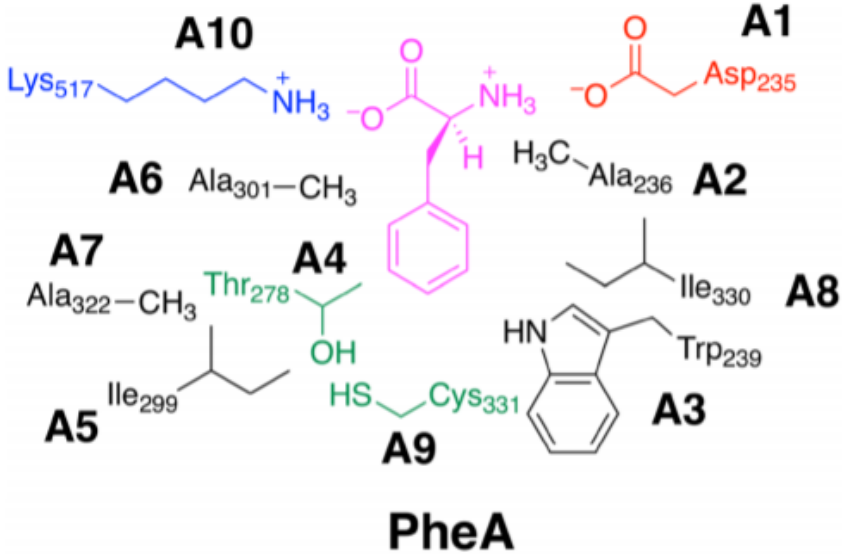


NRP-PK

(From : https://www.scripps.edu/shen/researchoverview/research2.html)

Engineering adenylation domain of NRPS

- Mechanism of substrate recognition of A domain have been extensively studied (Kudo et al., 2018).
  - A1(D235) and A10(K517) recognize the amino acids, including D-α-AA and β-AA
  - A2～A9 recognize the sidechains.
- It is now possible to predict the selectivity of the protein by genome sequence.
- Recently, some study on rational redesign or directed evolution of A-domain have been done.

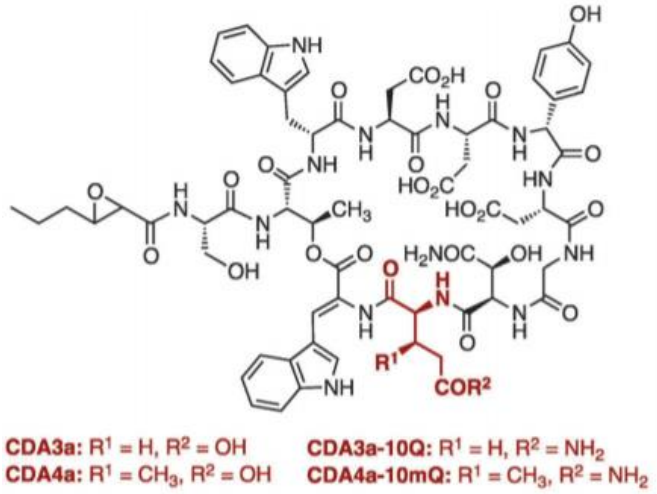


Kudo et al, *J. Ind. Microbiol. Biotechnol.*, **2018**, not yet assigned to an issue.
Stachelhaus et al., *Chem. Biol.*, **1999**, *6*, 493-505

Engineering adenylation domain of NRPS

Thirlway et al., 2012

- Rationally redesigned module 10 A-domain, CdaPS3, of calcium-dependent antibiotic to change the selectivity.
- Lys278Gln at A4 : L-Glu (or 3-methyl-L-Glu) ➡ 3-methyl-L-Gln
- By mimicking the other A-domains which recognize Gln.



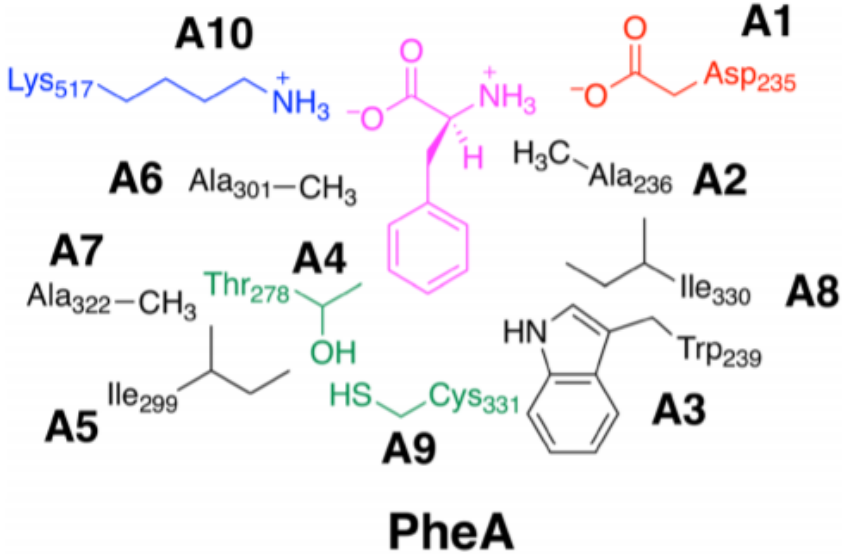
Protein	236	239	278	299	301	322	330	331
CdaPS3 (Glu)[a]	Gln	Gly	Lys	Thr	Gly	Val	Gly	His
SrfA (Glu)	Ala	Lys	Asp	Leu	Gly	Val	Val	Asp
FenA (Glu)	Ala	Trp	His	Phe	Gly	Ser	Val	Glu
LicA (Gln)	Ala	Gln	Asp	Leu	Gly	Val	Val	Asp
TycC (Gln)	Ala	Trp	Gln	Phe	Gly	Leu	Ile	Asp

Thirlway et al, *Angew. Chem. Int. Ed.*, **2012**, *51*. 7181-7184.

Engineering adenylation domain of NRPS

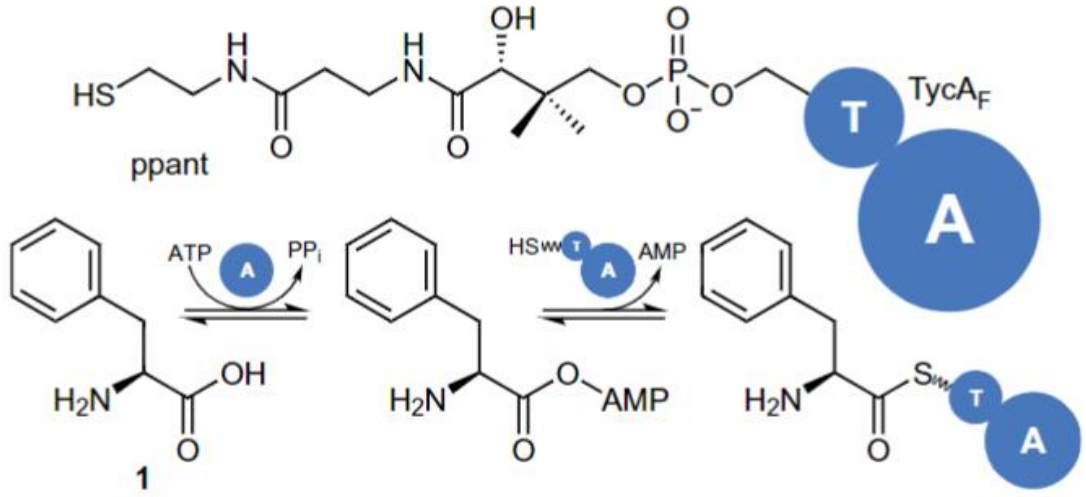
Kries et al., 2014

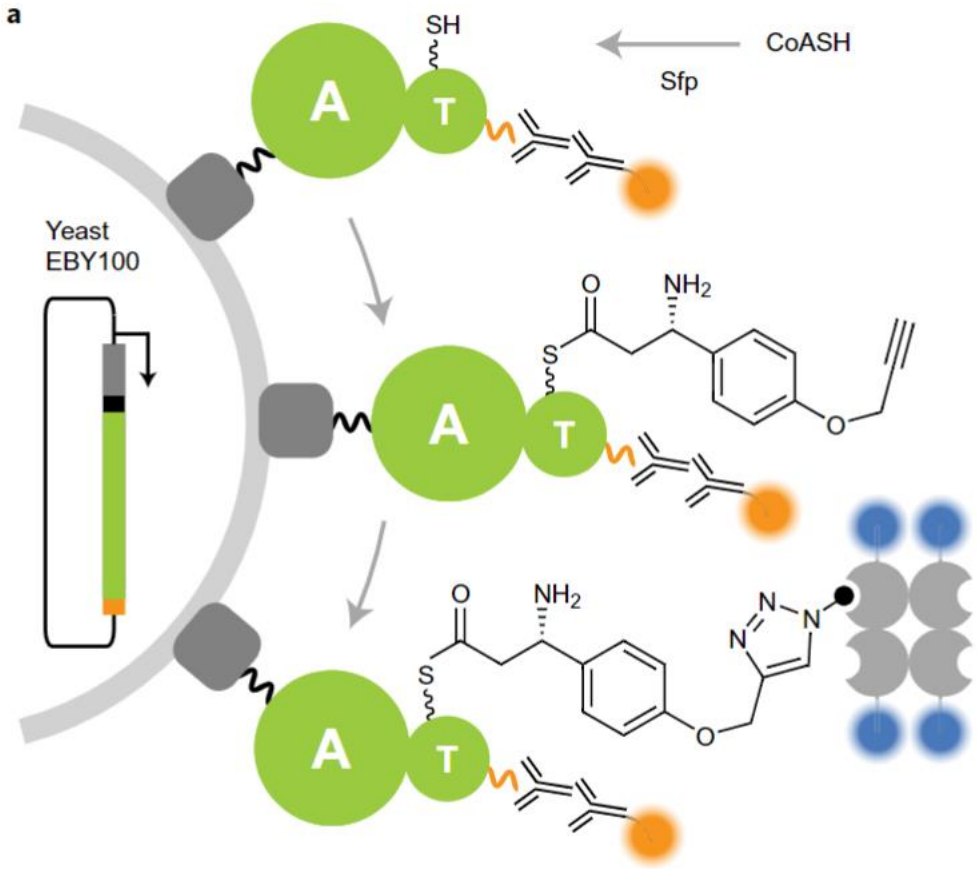
- Modified TycA (Phe specific) to recognize *O*-Propalgil Tyr
- Trp239Ser at A3 : L-Phe ➡ *O*-propalgyl-L-Tyr
- By altering A2～A9 by cassette mutagenesis
- This work is of importance since propalgil group can be a substrate of Huisgen Cycloaddition.



Engineering adenylation domain of NRPS

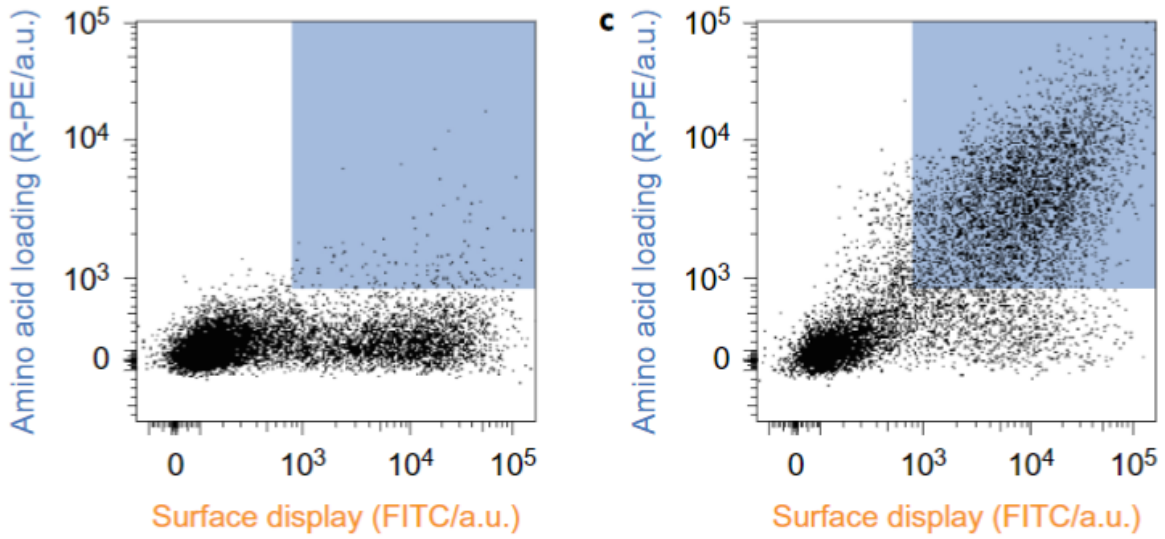
- Niquille et al., 2018 changed the specificity from α-amino acid to β-amino acid by high-throughput assay.
- $TycA_F$ was used as the target.

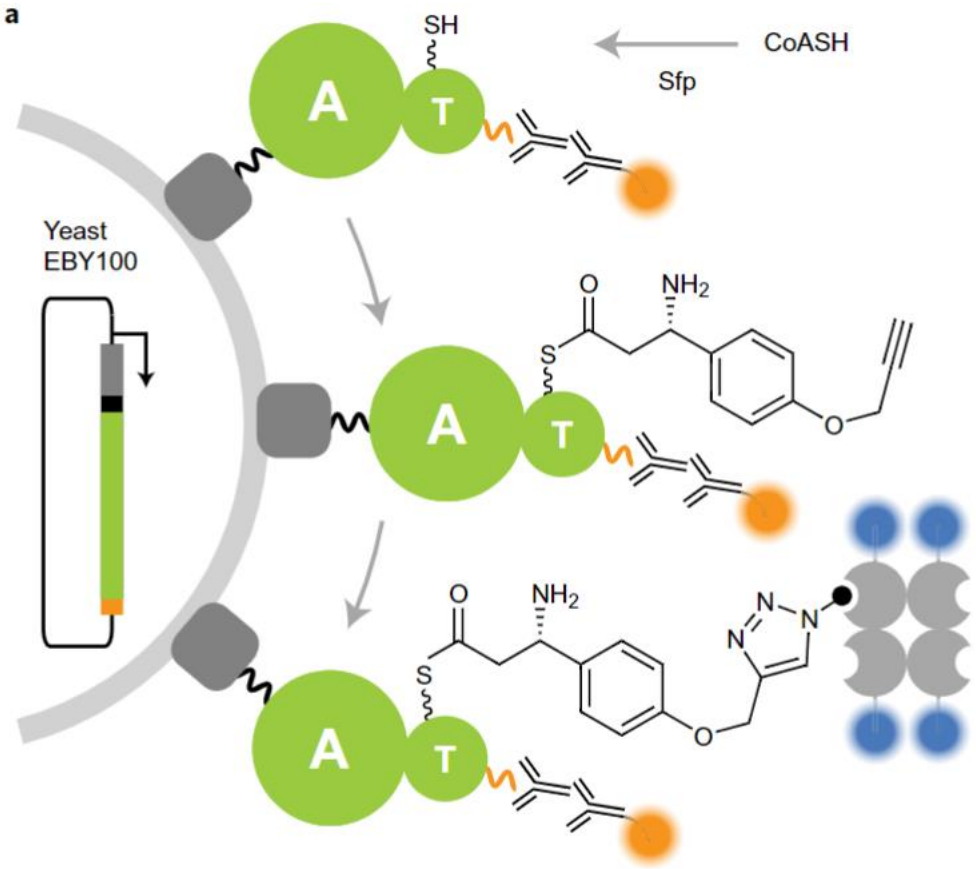




Engineering adenylation domain of NRPS

- In the paper, $TycA_{pY}$ was used to enable assay by fluorescence.
- Those strains were screened by FACS.

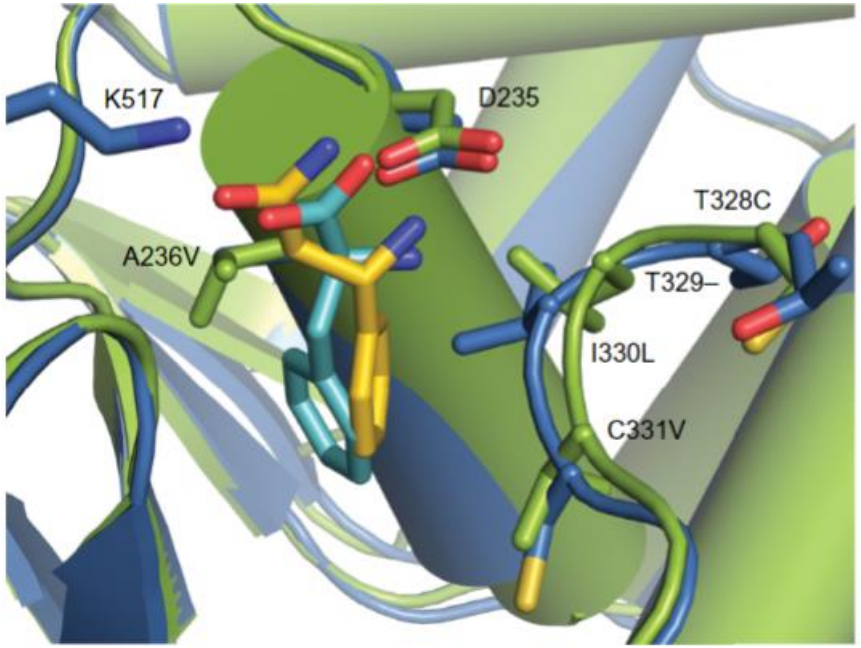




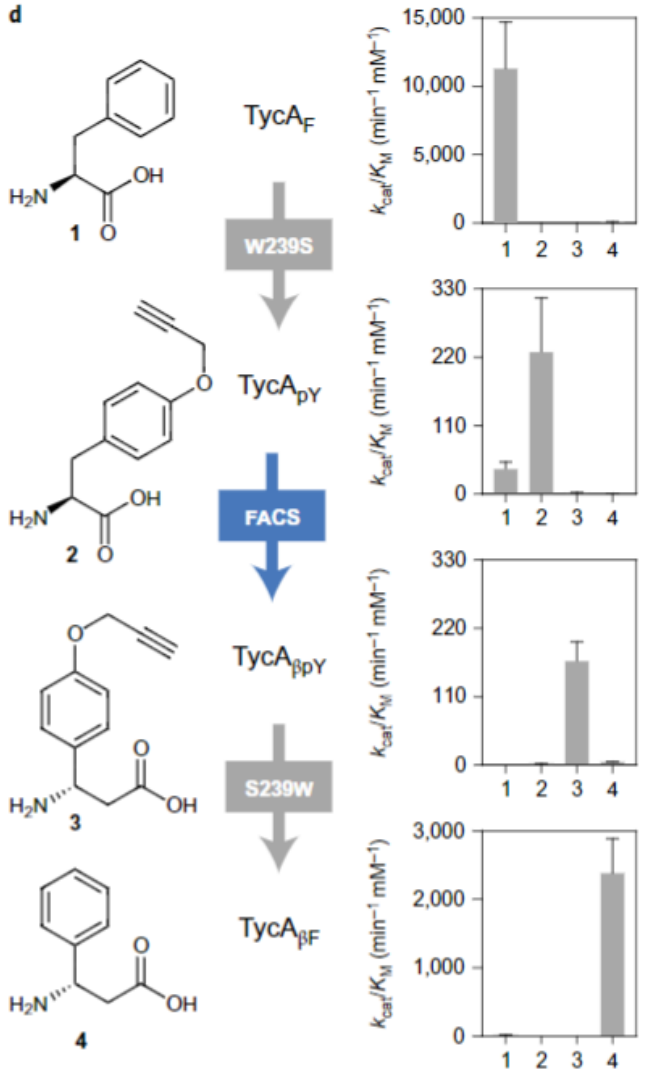
Niquille et al., *Nat. Chem.*, **2018**, *10*, 282-287

Engineering adenylation domain of NRPS

- Mutation was not fully random, but designed based on structural analysis.
- Reversion of W239S resulted in desired enzyme.

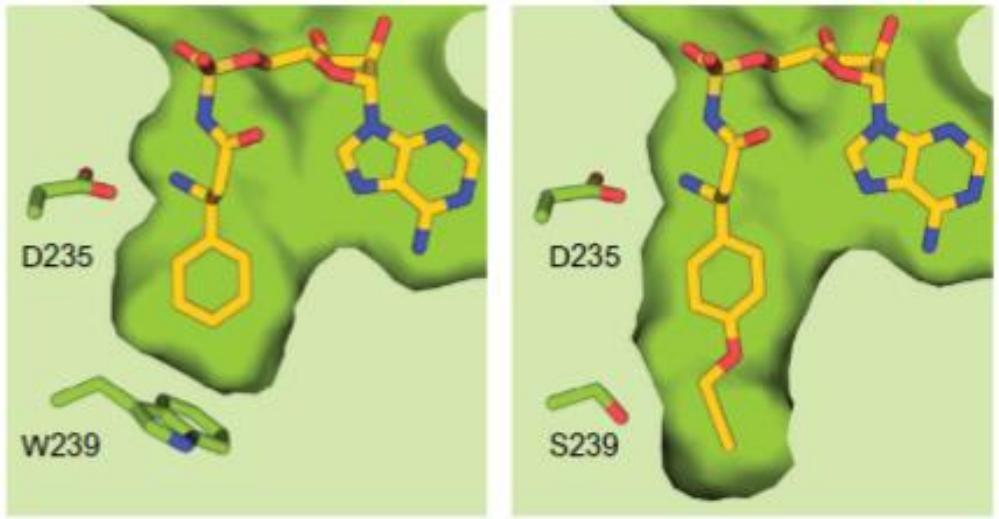


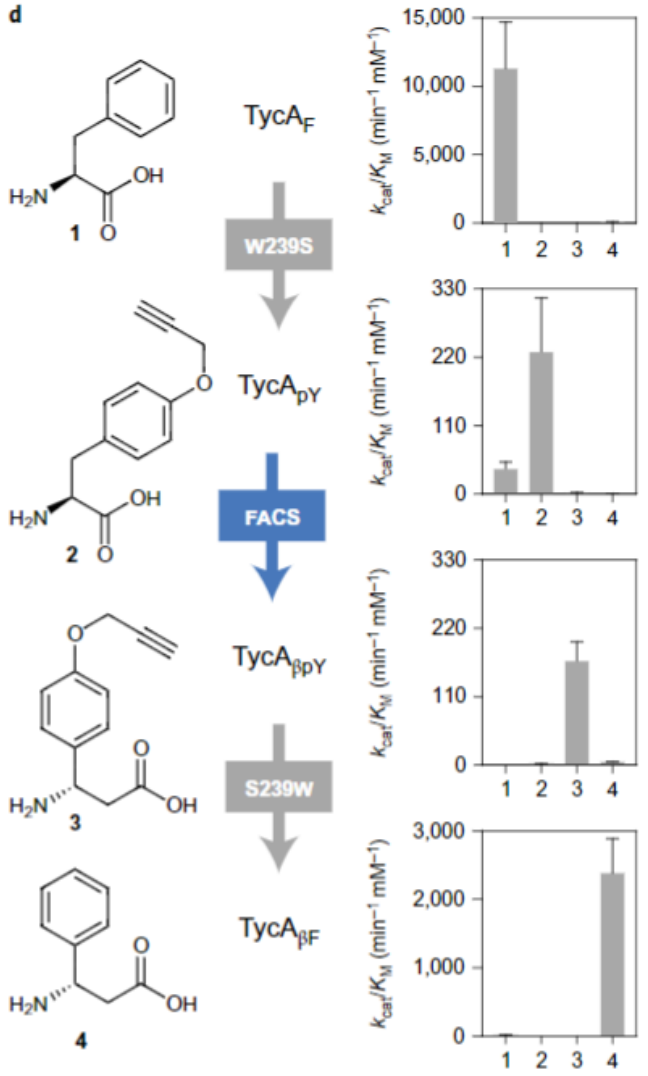
Superimposed structure of GrsA and $Tyr_{\beta F}$-A_N



Engineering adenylation domain of NRPS

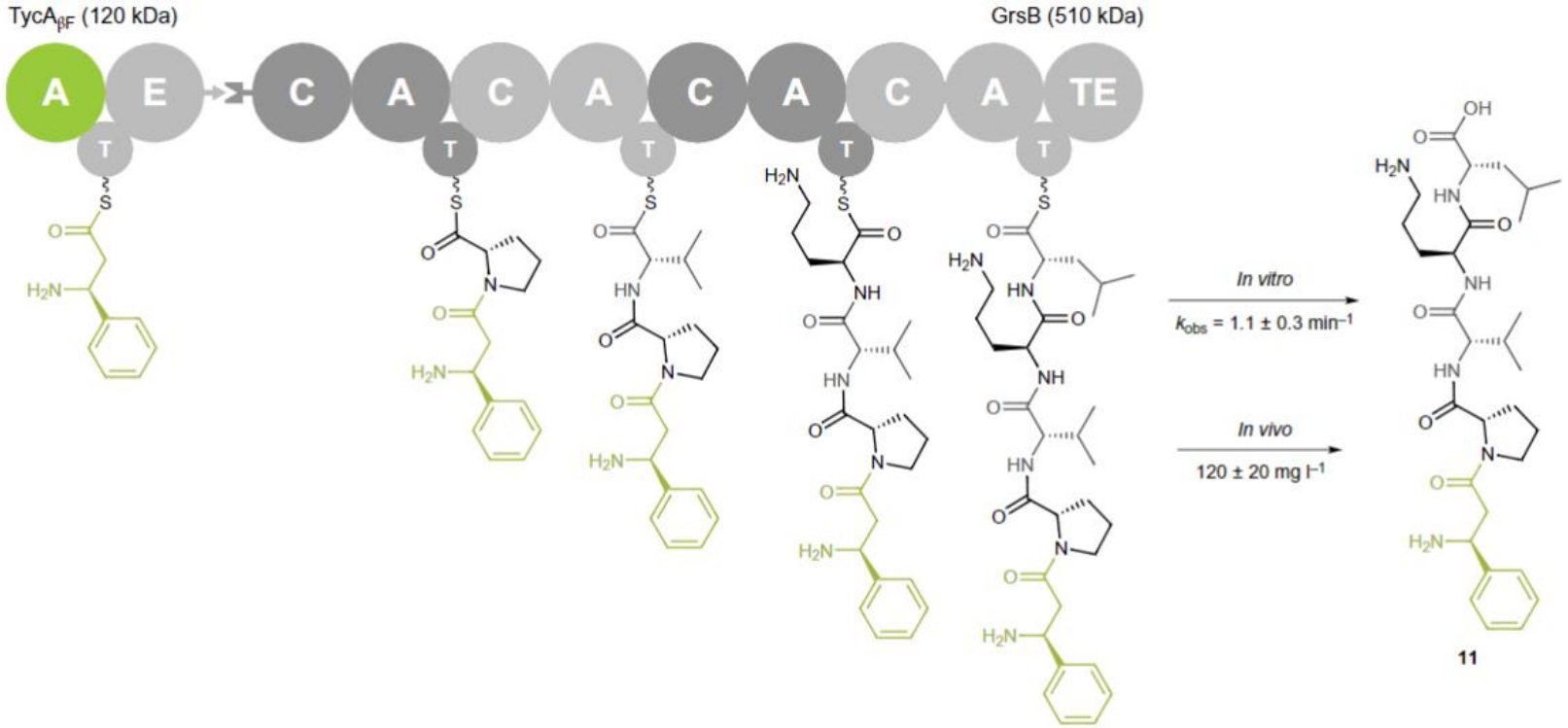
- Mutation was not fully random but designed based on structural analysis.
- Reversion of W239S resulted in desired enzyme.





Engineering adenylation domain of NRPS

- Authors confirmed that modified A domain correctly work in cooperation with other NRPSs.



Limitations

- Scope is limited to structurally conserved changes.
- The change in A domain specificity may result in unexpected effects on overall NRPS function.(Uguru et al., 2004)

1. Introduction
    - Secondary metabolites
2. Current progress of the field
    - Genome mining approach
    - Enzyme engineering approach
3. Summary

Appendix

Summary

- Diversity of natural products is worth exploiting for drug discovery, biological tool, and so on.
- However, discovery of brand-new NP is becoming difficult.
- To solve this problem, some approaches are possible, including
  - Epigenetically activating silenced BGC.
  - Engineering enzymes to alter the specificity.

Current issues and future

- Epigenetically activating silenced BGC.
  - Need Genome information, and establishment of transformation method.
  - Or, analysis of the product may be complicated.

➡The method without genetic transformation to activate specific gene cluster must be a breakthrough.

- Engineering enzymes to alter the specificity.
  - Despite the great effort, scope is limited to structurally conserved changes.

➡Using artificial catalyst to activate desired amino acid may be a solution.

  - The modification may result in unexpected overall effect.

1. Introduction
    - Secondary metabolites
2. Current progress of the field
    - Genome mining approach
    - Enzyme engineering approach
3. Summary

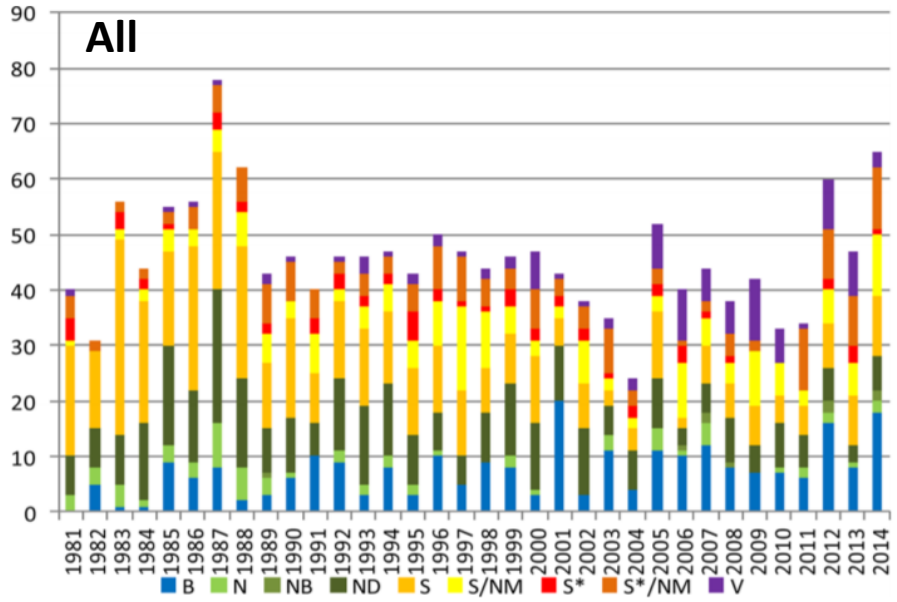
Appendix

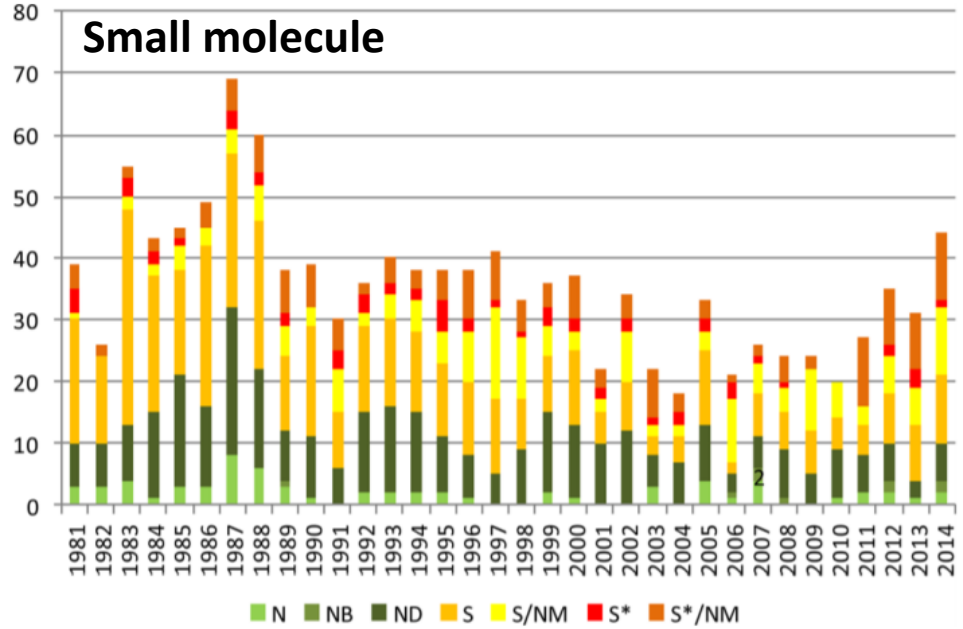
Appendix : The breakdown of all approved drug

B : Bio, V : Vaccine

N : NPs, NB : Botanical, ND : NP derivatives, S : synthetic, S* : Synthetic (NP pharmacophore)

/NM : Mimic of NPs





Newman and Cragg , *J. Nat. Prod.* **2016**, *79*, 629−661

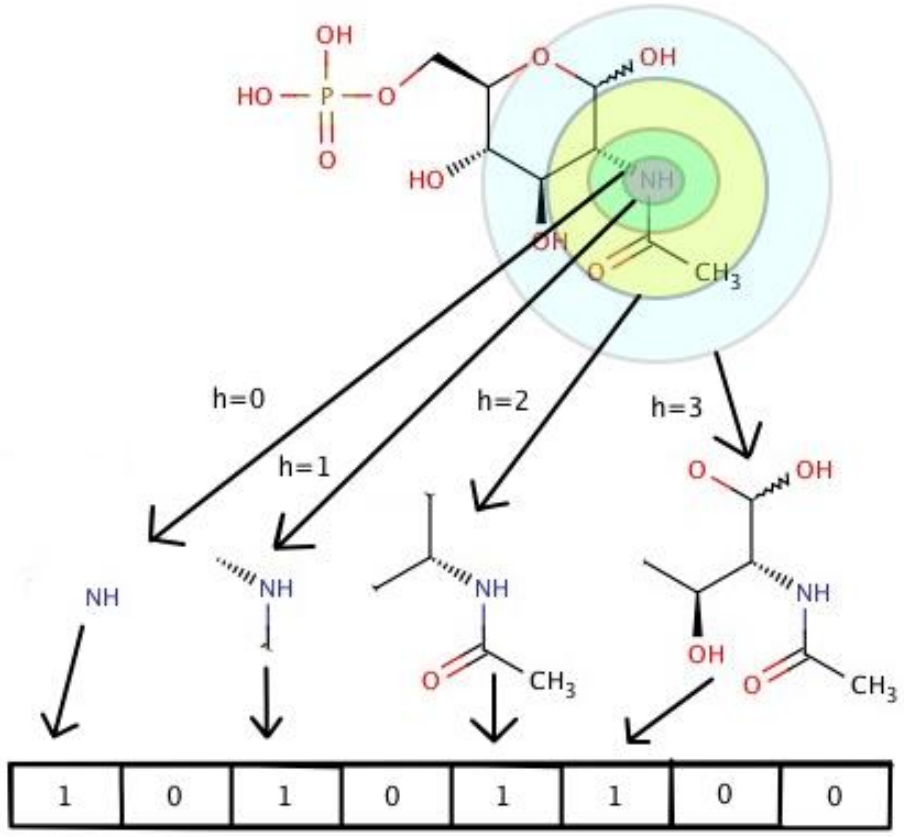
Appendix : Morgan fingerprint

Fingerprint calculated using Morgan Algorithm

(Extended-Connentivity Fingerprints)

For detail see the papers listed below.

Morgan, *J. Chem. Doc.*, ***1965***, *5*. 107-113.
Rogers and Hahn, *J. Chem. Inf. Model.*, **2010**, *50*, 742-754.



From : http://chembioinfo.com/2011/10/30/revisiting-molecular-hashed-fingerprints/

Appendix : Tanimoto scoring

Similarity score of two bitmaps, defined by right formula.

X_i : i th bit of X

$\wedge$ / $\vee$: bitwise operator And / Or.

$$T_s(X,Y) = \frac{\sum_i (X_i \wedge Y_i)}{\sum_i (X_i \vee Y_i)}$$

Appendix : Pleiotropic methods

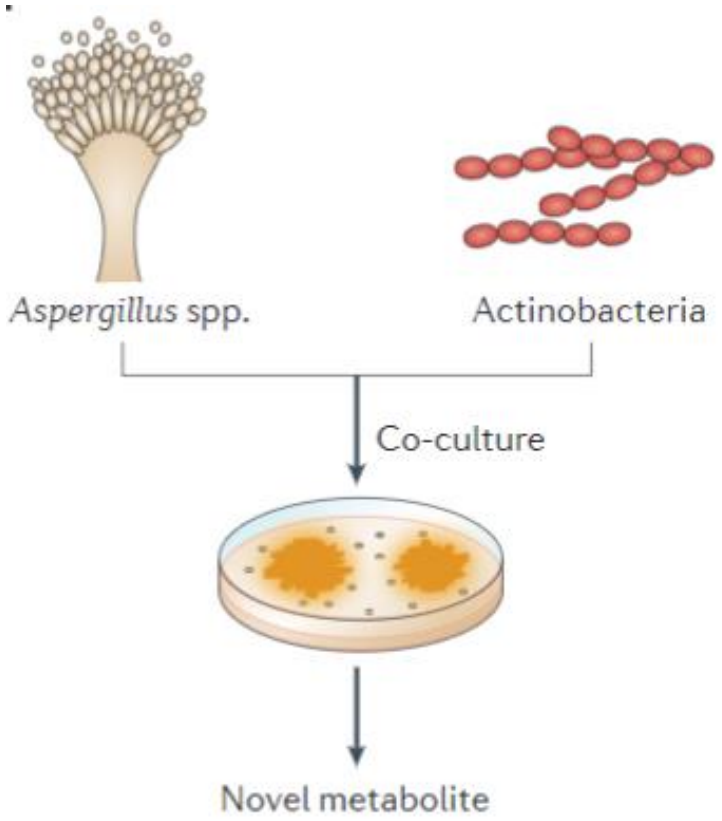
Variation in growth conditions

Approach : Just change the growth condition (co-culture, temperature, pH, etc.)

Advantage : Easy to conduct.

Limitations :
- Limited by the range of condition under which organism will grow.
- Difficult to predict the effect of the change.

e.g. : Aspoquinolons, Aspernidines, ClosthioamideA, etc.



Example of the approach.

Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.

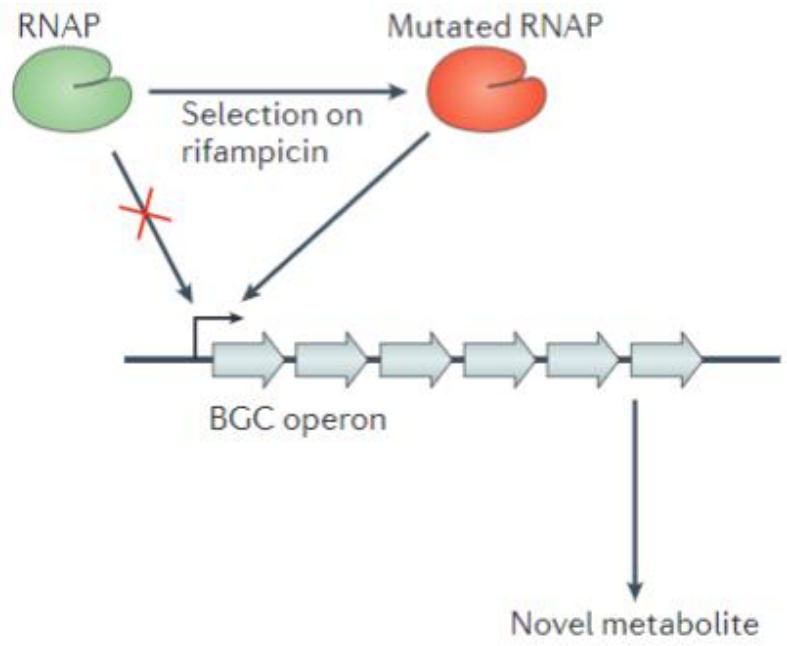
Appendix : Pleiotropic methods

Engineering the transcriptional and translational machinery

Approach : Induce mutation in RNA polymerase and ribosomal proteins to cause upregulation of BGC expression.

Limitations : Only small number of antibiotics trigger changes in a restricted group of bacteria.

e.g. : Piperidamycins, Coelimycins



Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.

Appendix : Pleiotropic methods

Manipulating global regulators

Approach : Change the expression levels of pleiotropic transcriptional regulator of BGCs.

Advantage : Manipulation of single regulator may result in more than one pathway and may lead to the discovery of multiple NPs.

Limitations : Only few such regulators are known.

e.g. : Terrequinone (*A.nidulans*), ClosthioamidesB-H, Pulvomycin (*S.flavopersicus*)

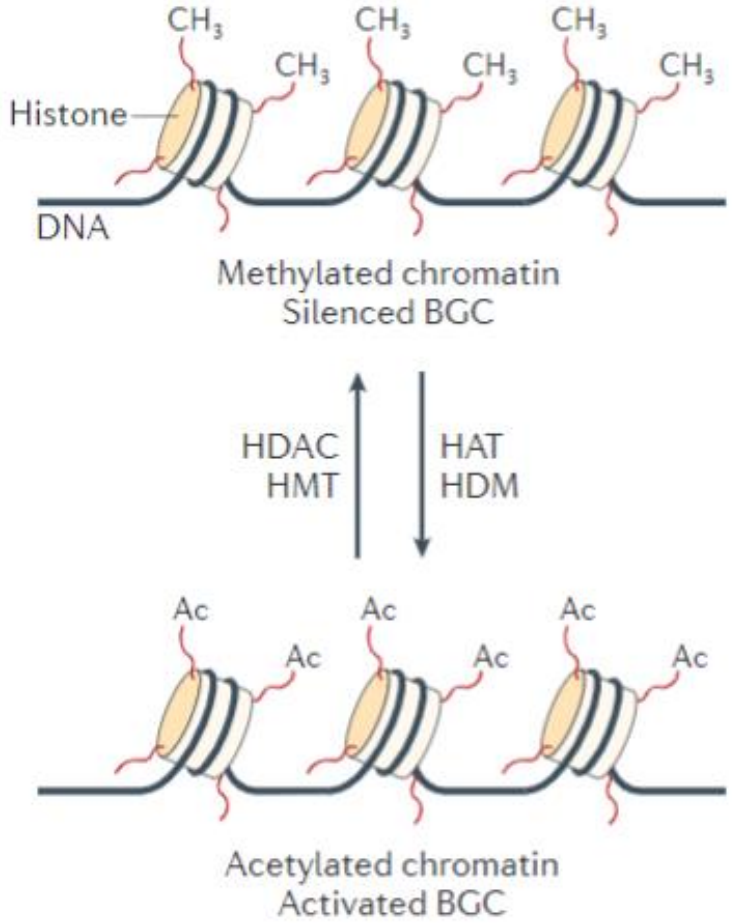
Appendix : Pleiotropic methods

Epigenetic perturbation

Approach : Use mutagenesis or inhibitors to trigger global chromatin structure change.

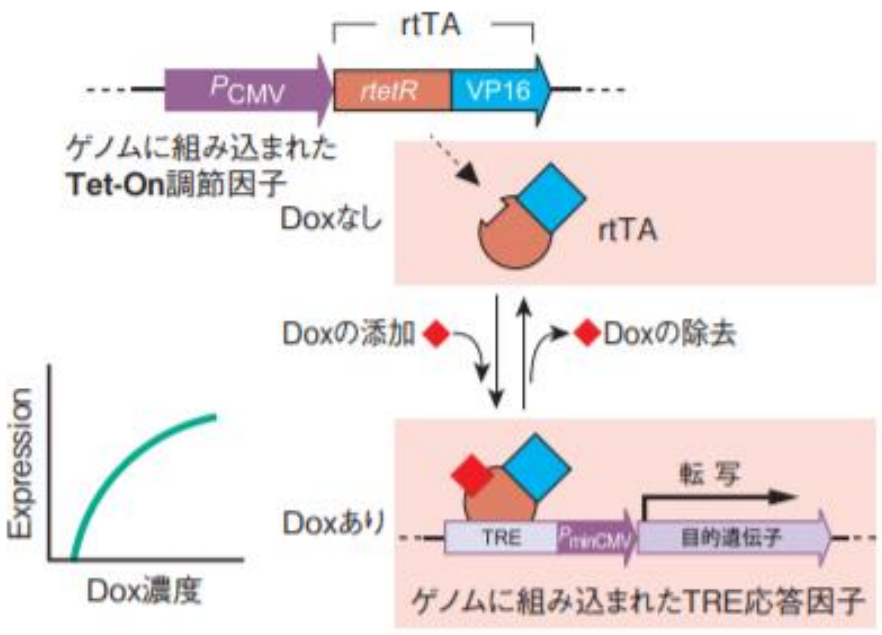
Limitations :
- Sites at which acetylation or methylation may be altered are limited by accessibility.
- Not yet scope for predictive change.

e.g. : Emodin (*A.nidulans*), Cladochromes, NygeroneA

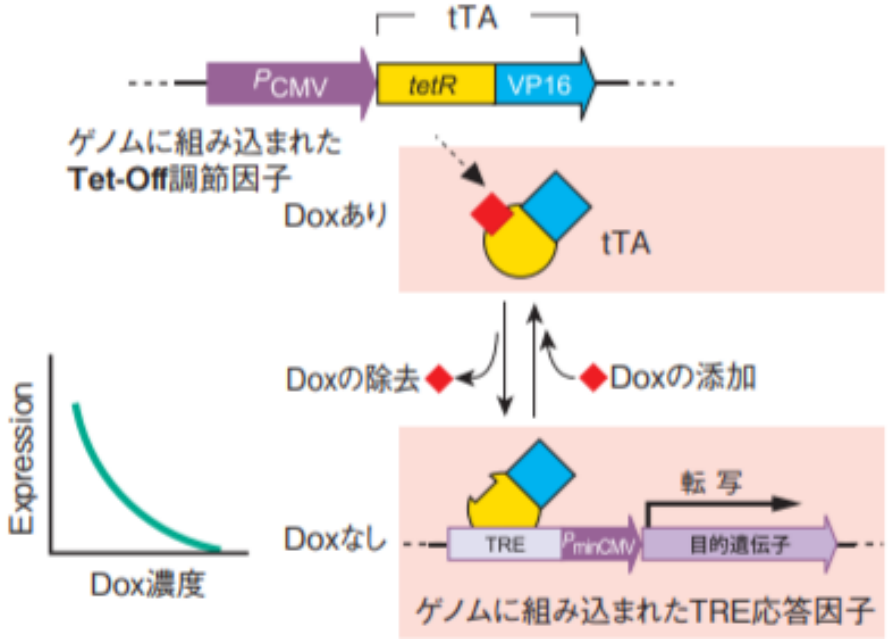


Appendix : Readily inducible promoter

- Tet-ON system



- Tet-OFF system



(From : http://www.takara-bio.co.jp/goods/catalog/pdf/tet.pdf)

Appendix : Pathway-specific methods

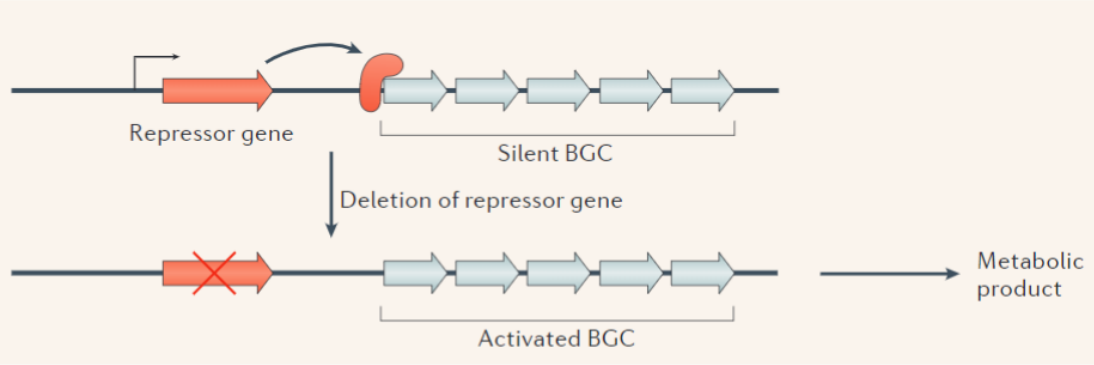
Manipulating pathway specific regulators

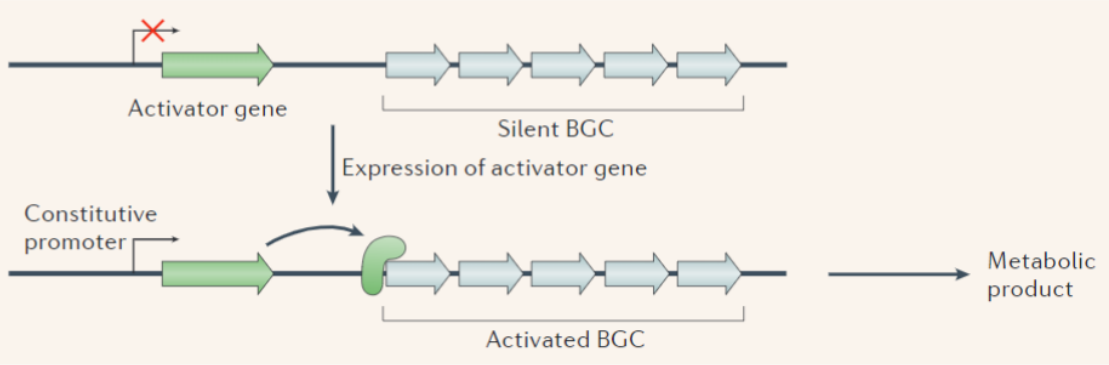
Approach : Identification and overexpression or deletion of pathway-specific regulators.

Advantage : Precise activation of specific BGC. Unambiguous identification of the products.

Limitations : Transcription factor should be identifiable, easy to manipulate and the perturbation should not complicate identification of the target metabolites.

e.g. : Aspyridones, Asperfuranone, Stambiomycins, Gaburedins, Burkholdacs, etc.





Appendix : Pathway-specific methods

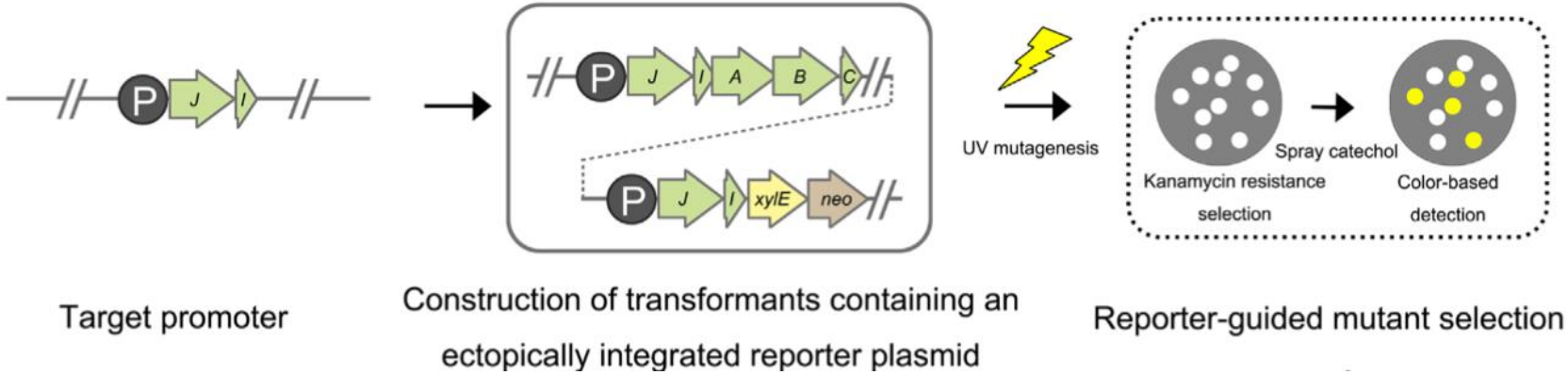
Reporter-guided mutant selection

Approach : Couple target BGC to promoter-reporter system, and induce genome scale random mutagenesis.

Advantage : Pleiotropic approach combined with pathway-specific detection of visualizing wonted mutant.

Limitations : Require cloning of promoter sequences. Only single example.

e.g. : Gaudimycin D, E



Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.
Guo, F et al. *Metab. Eng.*, **2015**, *28*, 134-142

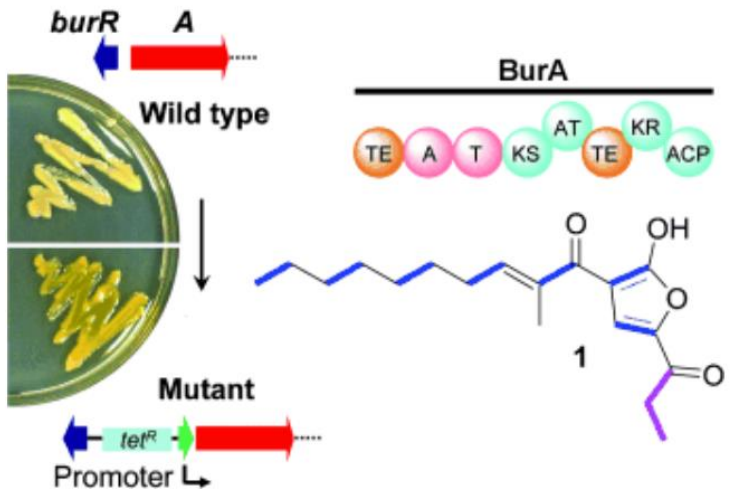
Appendix : Pathway-specific methods

Refactoring

Approach : Replace the promoter of the silent BGC with constitutive, or readily inducible promoter .

Advantage : Afford precisely controlled activation of the pathway of interest.

Limitations :
- Natural promoter must be identified.
- The BGC must be amenable to genetic manipulation



e.g. :Burkholderic acid, 6-epialteramideA/B

Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523.
Franke et al., *Angew. Chem. Int. Ed. Engl.*, **2012**, *51*, 11611-11615

Appendix : Pathway-specific methods

Heterologous expression

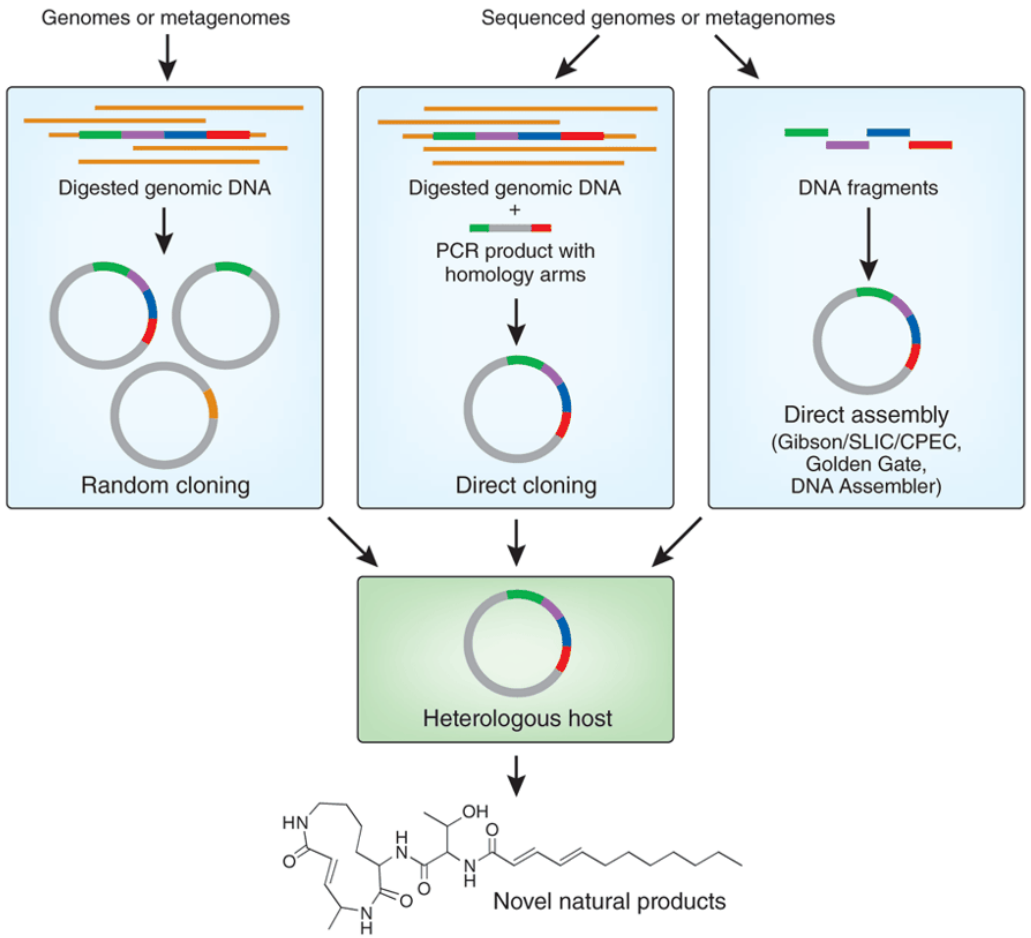
Approach : ▪Expression of an entire BGC in a heterologous host

▪Often combined with refactoring.

Advantage : SImplifies the identification of metabolites.

Limitations : BGCs are usually difficult to handle because of the size of the cluster.

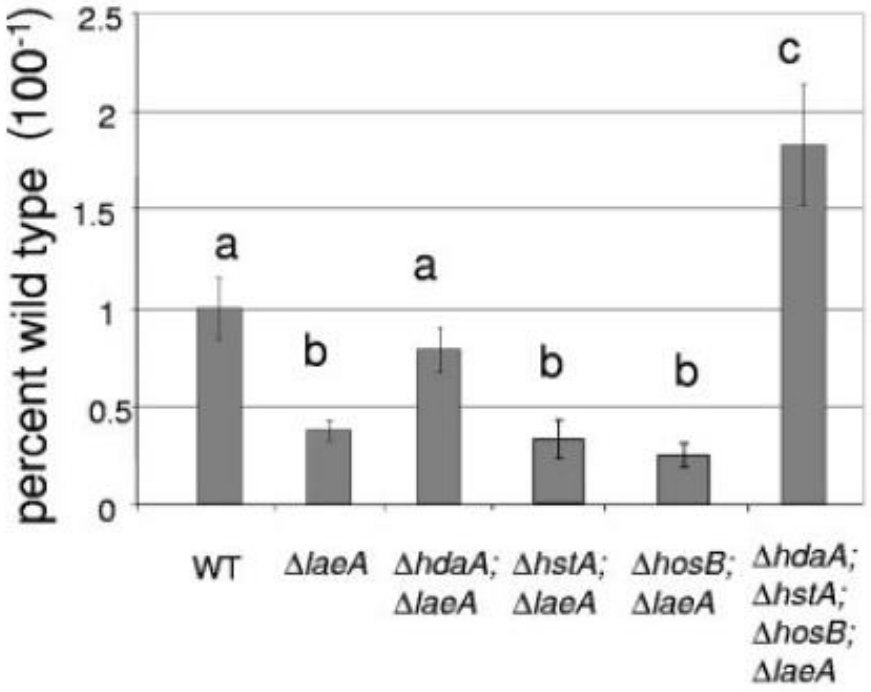
e.g. : Terferol analogues, Epiisozizaene, Acermitilol, Haloduracin



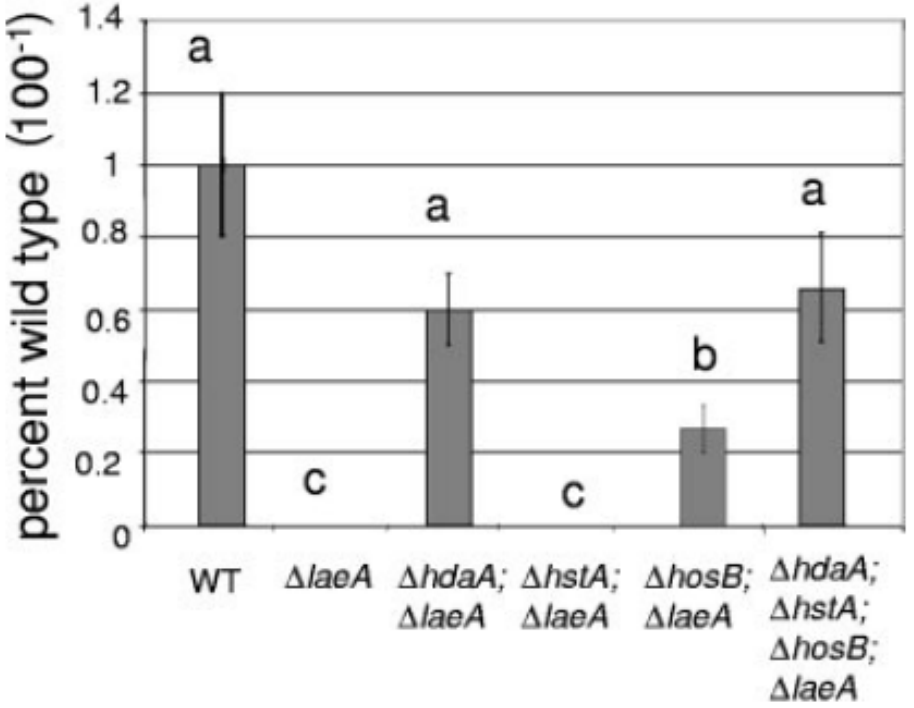
Rutledge and Challis, *Nat. Rev.*, **2015,** *13*, 509-523
Cobb and Zhao, *Nat. Biotech.*, **2012**, *30*, 405–406

Appendix : LeaA

NOR

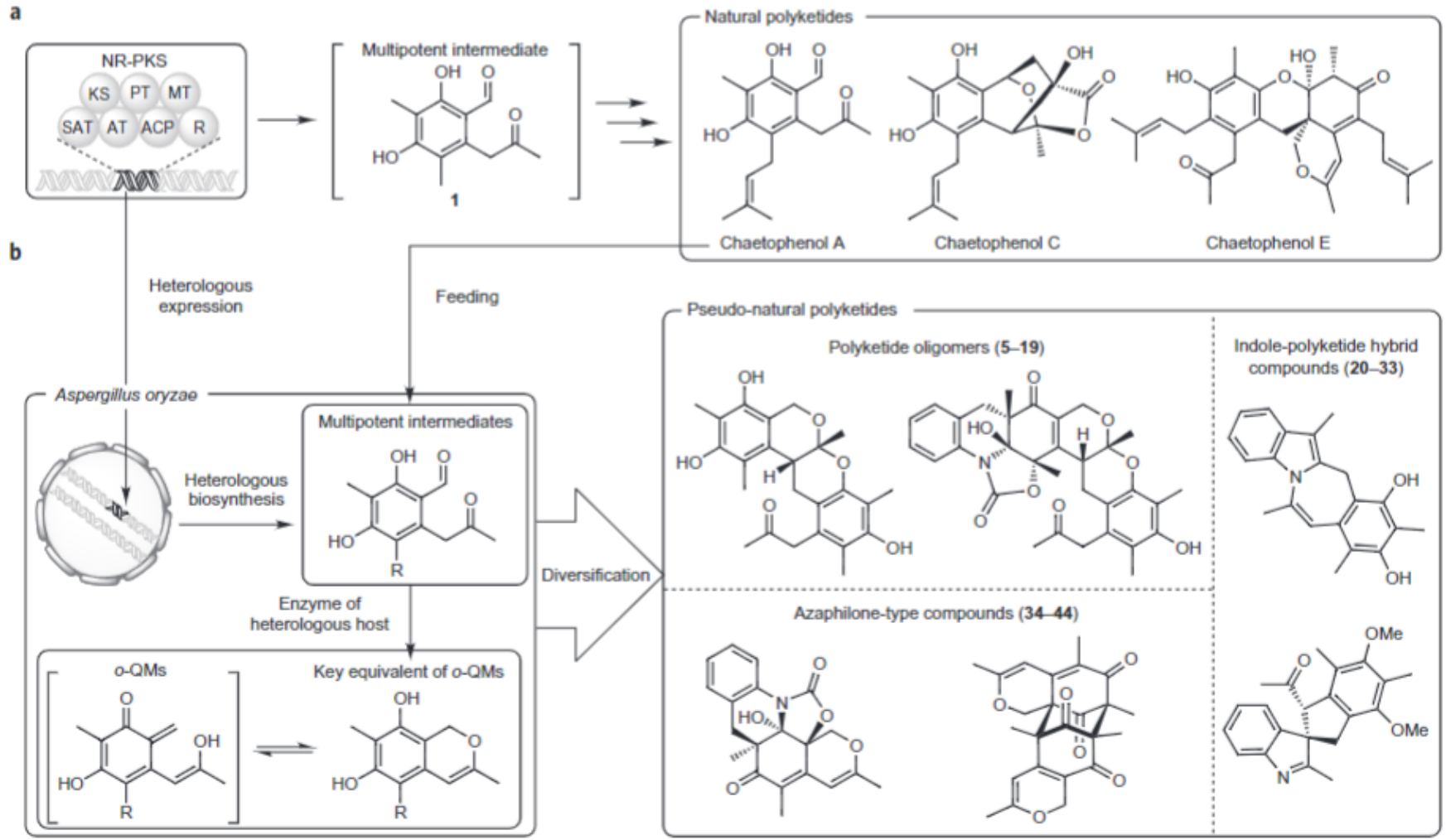


PN



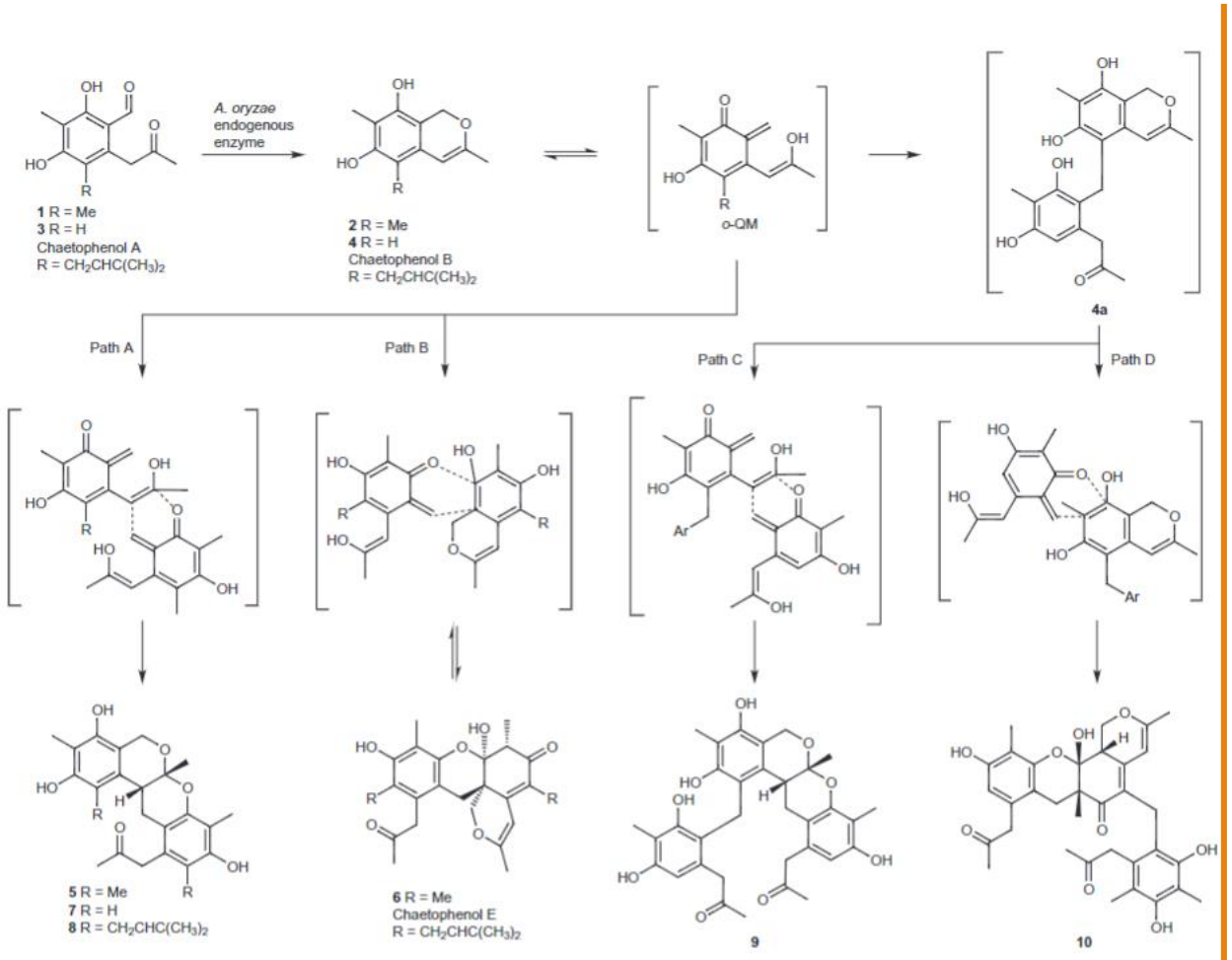
Shwab et al., *Eukaryotic Cell*, **2007**, *6*, 1656-1664

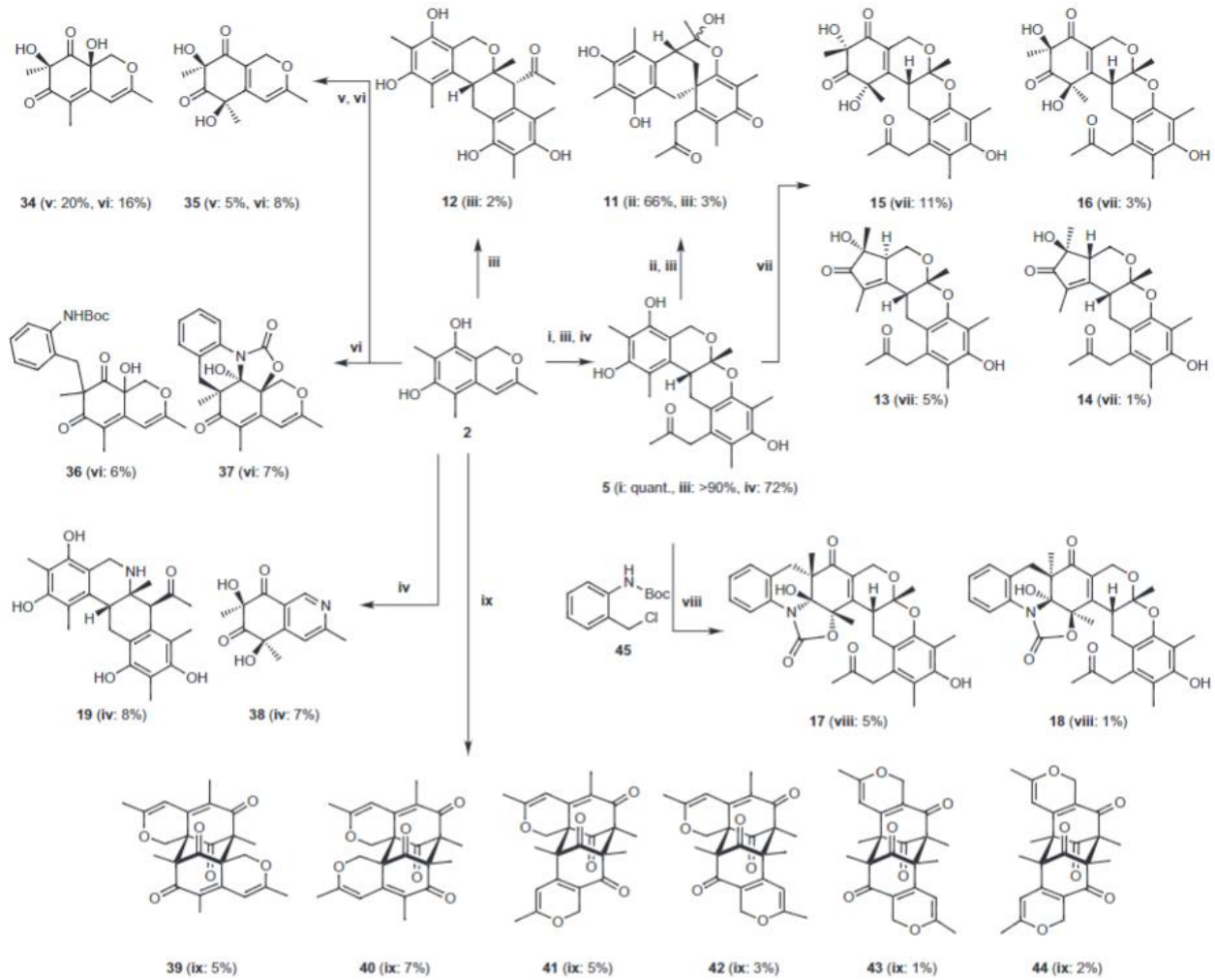
Appendix : Asai et al. 2015



Asai et al., *Nat. Chem.*,**2015**, *7*, 737-743

Appendix : Asai et al. 2015





Appendix : A domain structural analysis

Table 1 Structural analysis of amino acid adenylation enzymes

Amino acid substrate	Protein	Final natural products	Ligands, note	PDB ID[b]	Year	Reference
Proteinogenic L-α-amino acids						
L-α-Phe	PheA in GrsA	Gramicidin	L-α-phenylalanine and AMP	1AMU	1997	[19]
L-α-Phe	McyG	Microcystin	L-Phe-AMP, A-CP didomain	4R0M	2015	[99]
L-α-Arg/Tyr	ApnA-A1	Anabenopeptin	Arg/Tyr adenylate	4D57/56	2015	[48]
L-α-Leu	SrfA-C	Surfactin	None, C-A-T-TE full module	2VSQ	2008	[101]
L-α-Val	PA1211	Unknown	Valine–adenosine vinyl-sulfonamide, A-CP	4DG8, 4DG9	2012	[72]
L-α-Val	LgrA	Linear gramicidin	Val-NH-phospho-pantetheine attached to the PCP domain, F-A-CP initiation module	5ES8	2016	[87]
Gly	AlmE	LPS modification	Gly-AMP	4OXI	2014	[38]
Gly	AB3404	Unknown	Gly and AMP, C-A-T-TE full module	4ZXI	2016	[23]
L-Ser	EntF	Enterobactin	Ser-AVS, C-A-T-TE full module and with MbtH	5T3D	2016	[23][71]
L-α-Thr	Thr1	Chloro Thr	Thr and ATP, Thr-AMP	5N9 W, 5N9X	2017	[90]
Nonproteinogenic amino acids						
D-α-amino acids						
D-Ala	DltA	D-alanylation of lipoteichoic acid	D-Ala-AMP	3DHV	2008	[25][26] [114]
Modified α-amino acids:						
(2*S*,3*S*)-β-methyl-L-aspartic acid	SlgN1	Streptolydigin	AMP-cpp, as α-amino acids	4GR5	2013	[39]
N-cis-anhydromevalonyl-*N*-hydroxy-L-ornithine	SidNA3	Fungal siderophore	Fungal adenylation enzyme	3ITE	2010	[56]
Norcoronamic acid (and L-Val)[a]	ThioS($A_{4a}M_4A_{4b}$)	Thiocoraline	MT domain insertion type	5WMM	2018	[77]
2-aminobutyric acid	CytC1	Cytotrienin	Only pdb data. L-valine is the best substrate	3VNR, 3VNQ, 3VNS	2007	[105]
β-amino acids:						
(2*S*,3*S*)-β-methyl-L-aspartic acid	VinN	Vicenistatin	β-methyl-L-aspartic acid	3WV5	2014	[73]
(3*S*)-3-aminobutyric acid	IdnL1	Incednine	3-ABA-adenylate	5JJQ	2017	[17]
3-aminononanoic acid	CmiS6	Cremimycin	None	5JJP	2017	[17]
(*S*)-β-phenylalanine	engineered TycA	–	β-Phe-AMS (sulfamoyl adenosine)	5N82	2018	[80]
Others						
Anthranilic acid	AuaEII	Aurachin	Anthraniloyl-AMP, CoA ligase	4WV3	2016	[43]

Appendix : fluorescence activated cell sorting (FACS)

- Cells are arranged in line, and analyzed by laser.
- Drop containing target particle which satisfy the criteria is charged.
- The drops are deflected by charged plate, and collected.