

实用生物信息技术期末总结交流报告会  
**Semester Summary Seminar for Applied Bioinformatics Course**

次级代谢产物及其生物合成基因簇  
**Secondary Metabolites and  
their Biosynthesis Gene Clusters**

报告人 段佳琪  
中国农业科学院研究生院  
2020级博士班

2020年12月12日

1. Secondary Metabolites
2. Biosynthesis Gene Clusters
3. Prediction
4. Outlook

# 1. 次级代谢产物 (Secondary Metabolites, SMs)

某种细胞或者某种生物体的特有代谢途径中产生的小分子化合物。



红豆杉



黄花蒿

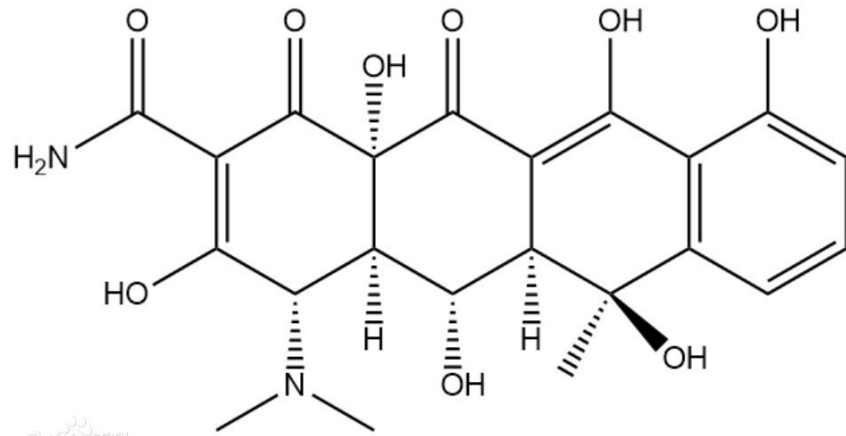


人参

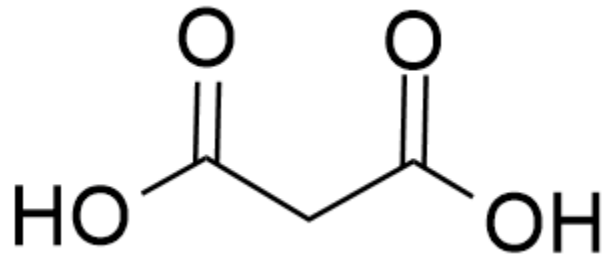
图片来源：百度百科

# 1.1 聚酮类化合物 (Polyketones, PKs)

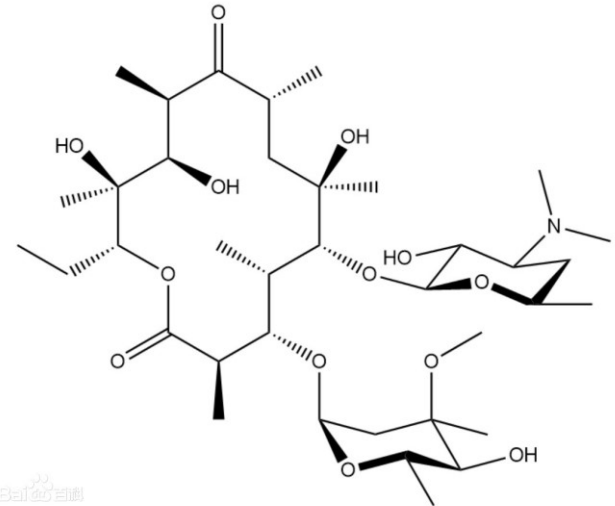
- 土霉素



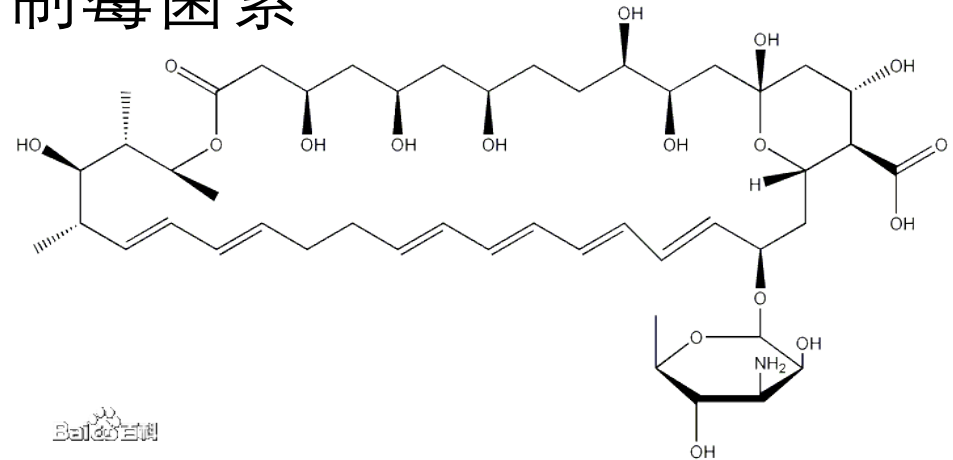
- 基本合成单元



- 红霉素

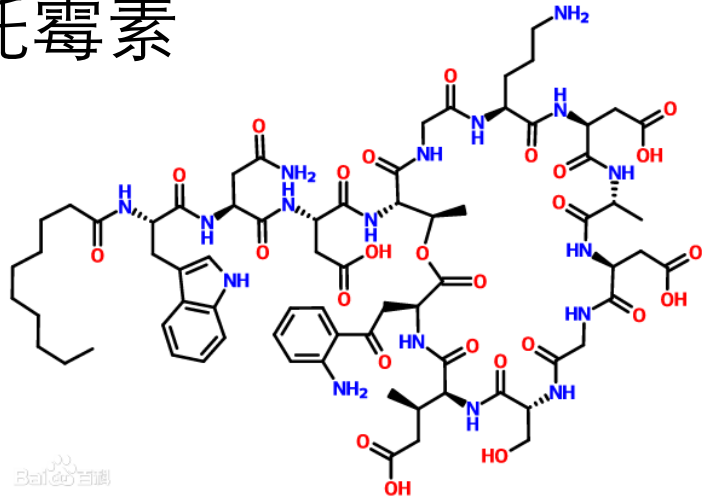


- 制霉菌素

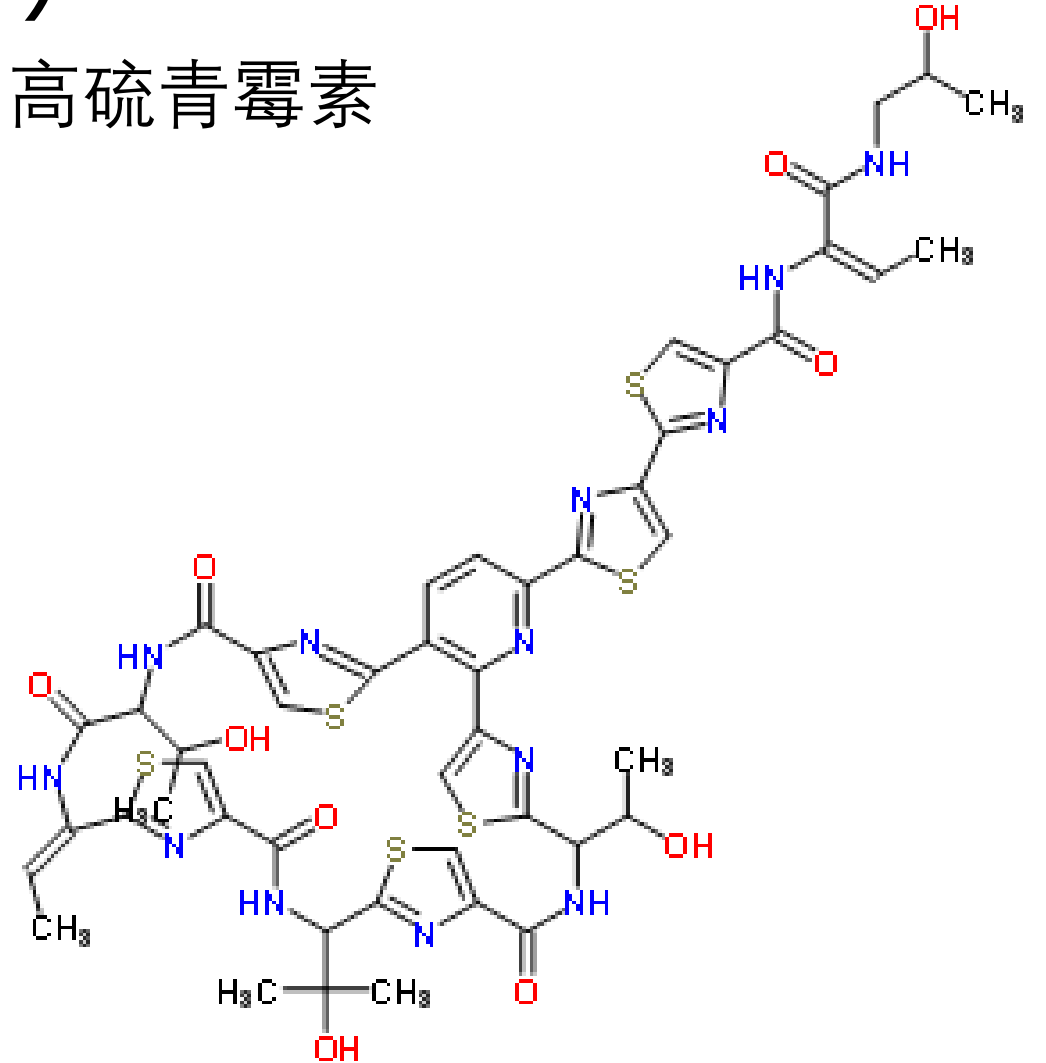


# 1.2 肽类化合物 (Petides)

- 达托霉素



- 高硫青霉素

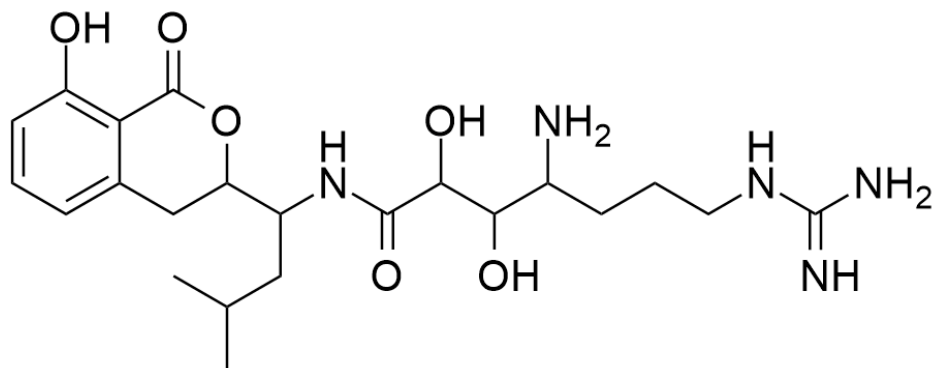


- 基本合成单元

天然氨基酸与非天然氨基酸

# 1.3 NRP-PK杂合化合物

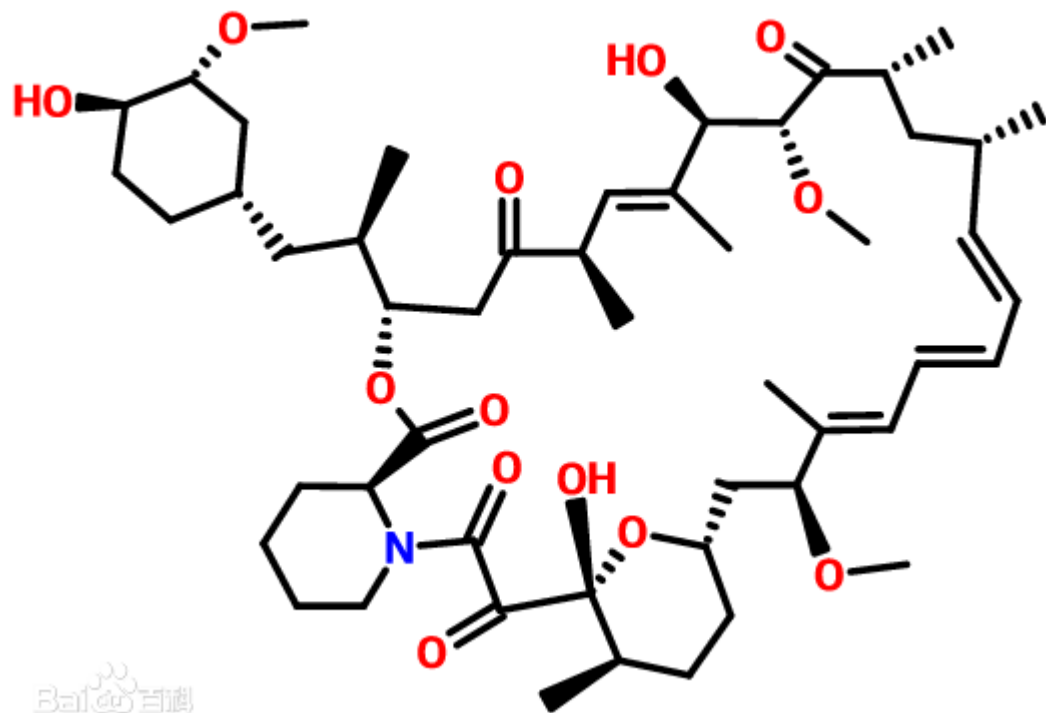
- Xenocoumacin, Xcn1



- 基本合成单元

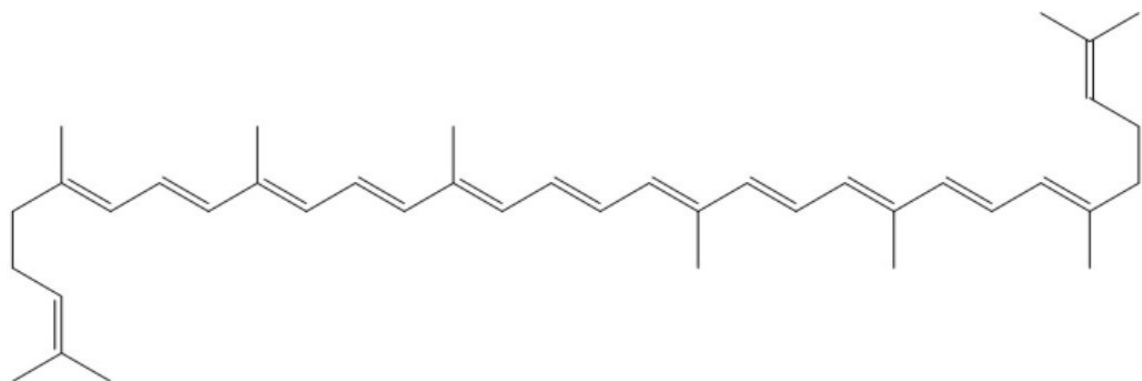
丙二酸、氨基酸等

- 雷帕霉素



# 1.4 异戊二烯类/萜类化合物

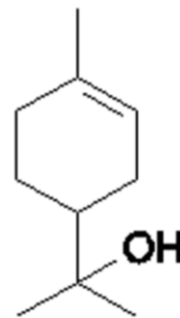
- 番茄红素



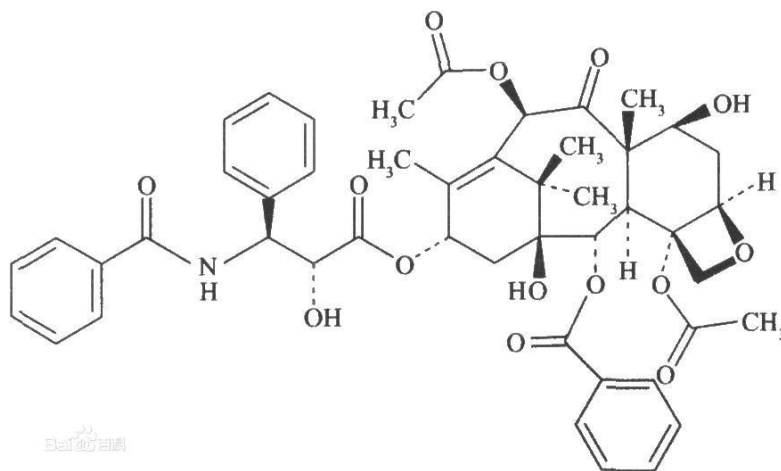
樟脑



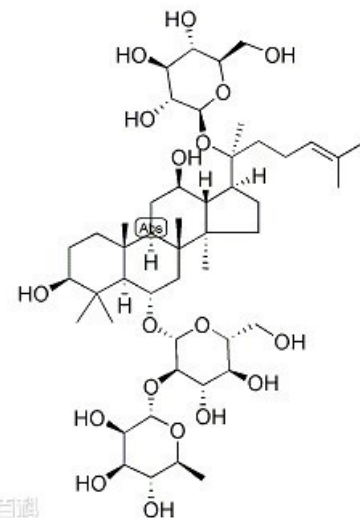
松油醇



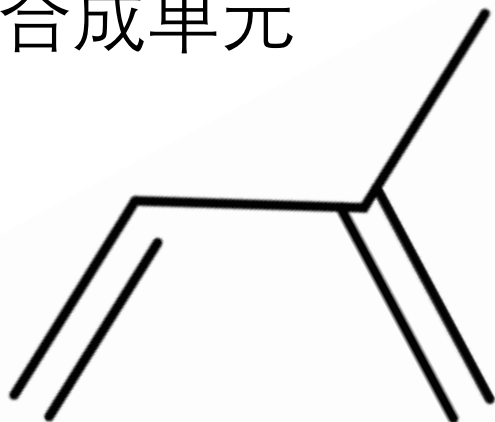
紫杉醇



人参皂苷



- 基本合成单元

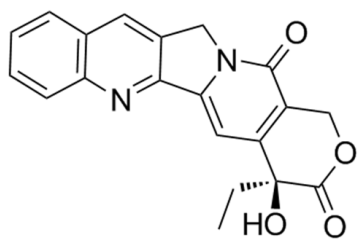


图片来源：百度百科

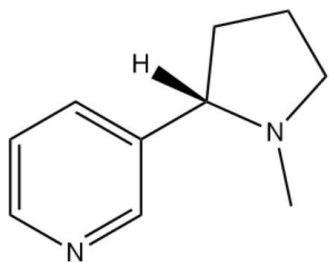
图片来源：百度百科

# 1.5 生物碱及其它化合物

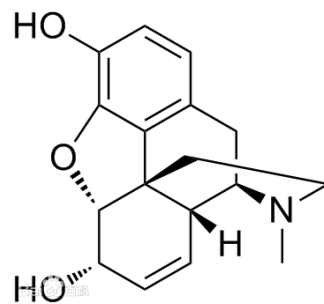
喜树碱



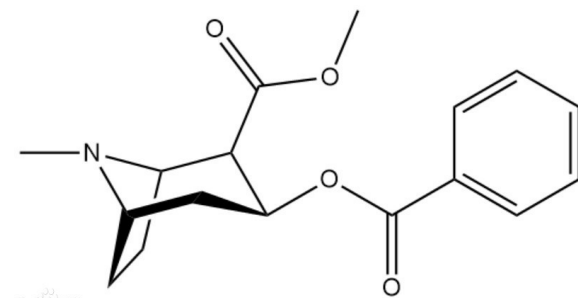
尼古丁



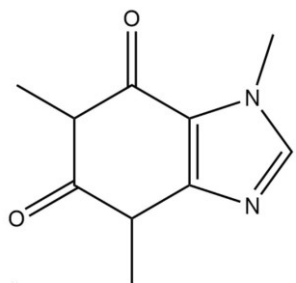
吗啡



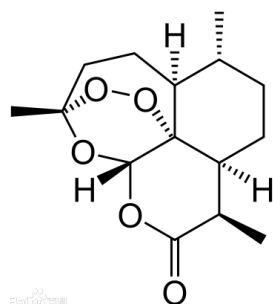
可卡因



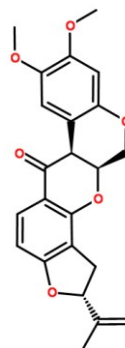
咖啡因



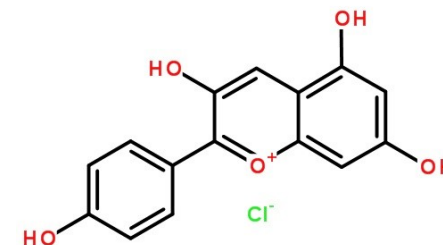
青蒿素



鱼藤酮



花青素

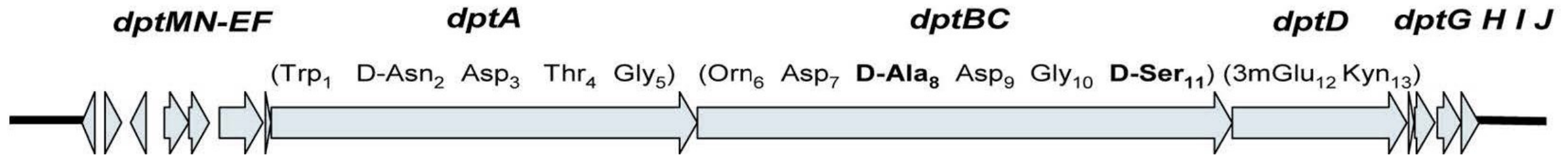


图片来源：百度百科



## 2. 生物合成基因簇 (Biosynthesis Gene Clusters, BGCs)

生物基因组中紧密相连，成簇存在，在功能上相互协同，形成合成次级代谢产物完整途径的一系列基因。



达托霉素生物合成基因簇 (Kien T. Nguyen, PNAS, 2006)

## 2.1 聚酮类化合物生物合成方式

聚酮类化合物的分子骨架主要由生物合成基因簇中的聚酮合酶(polyketide synthase, PKS)合成。PKS是一类具有多种功能域的大型合酶, 该酶作用方式类似于脂肪酸合酶(fat acid synthase, FAS)。

**AT:** acyltransferase

**ACP:** acyl carrier protein

**TE:** thioesterase

**KS:** keto-acyl synthase

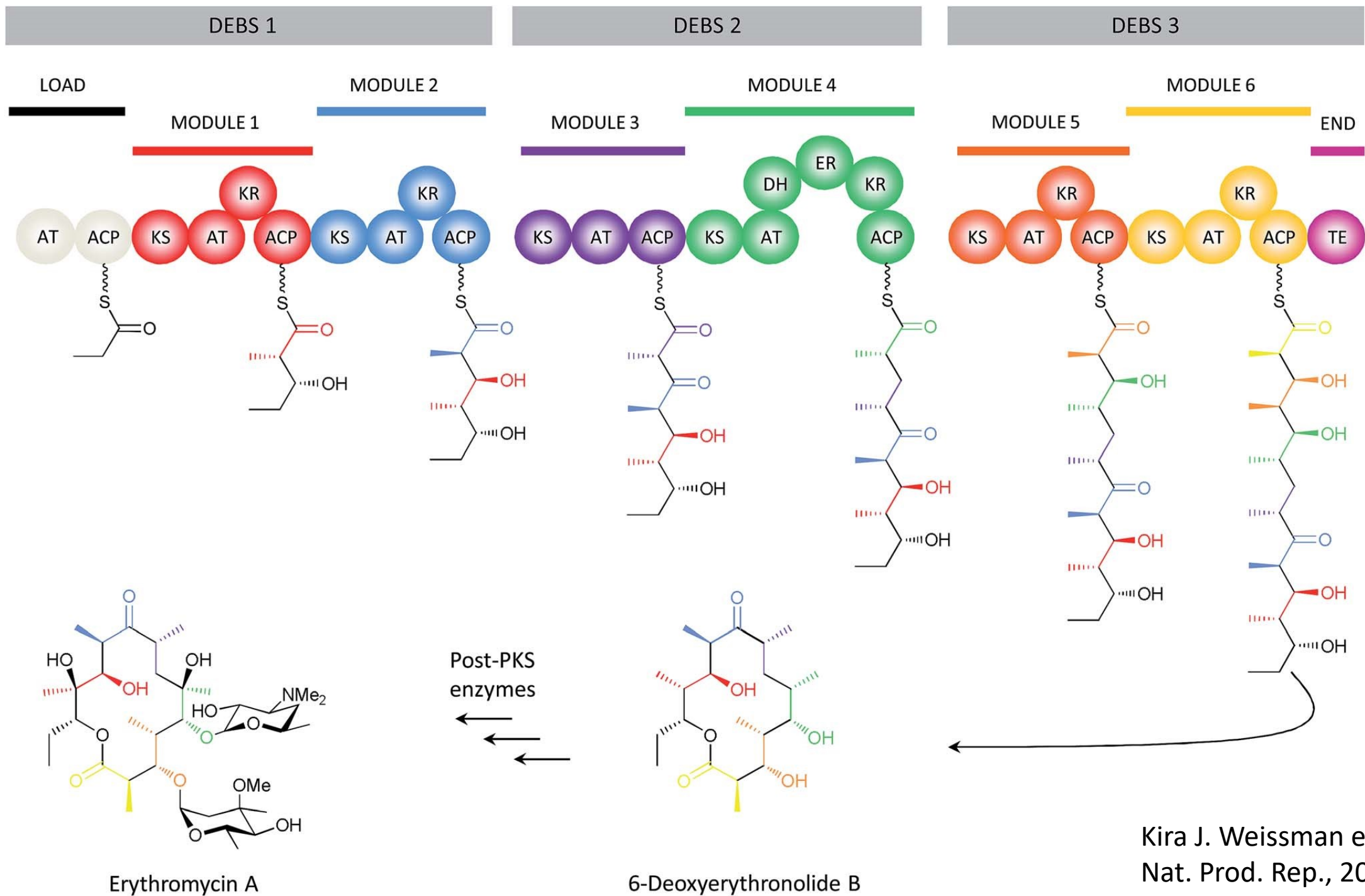
**KR:** ketoacyl reductase

**DH:** dehydratase

**ER:** enoylreductase

**MT:** methyl transferase

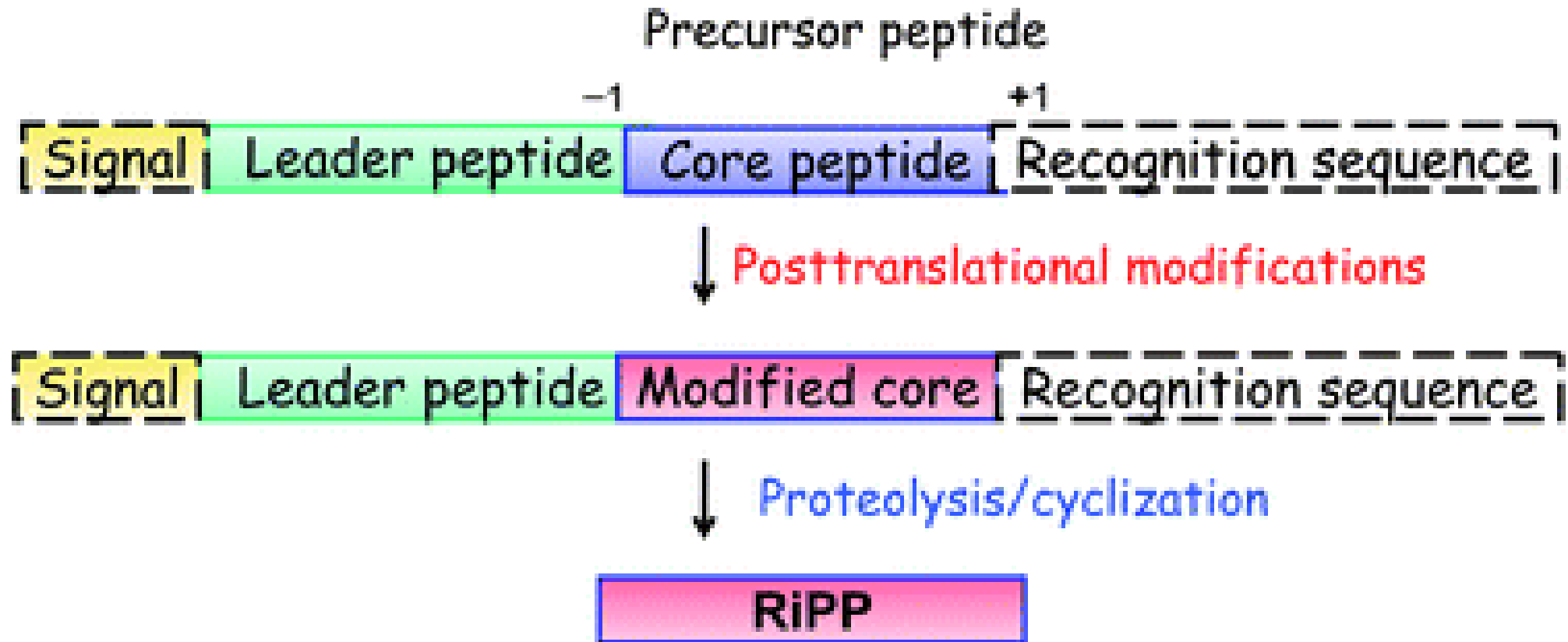
**Cy:** cyclase



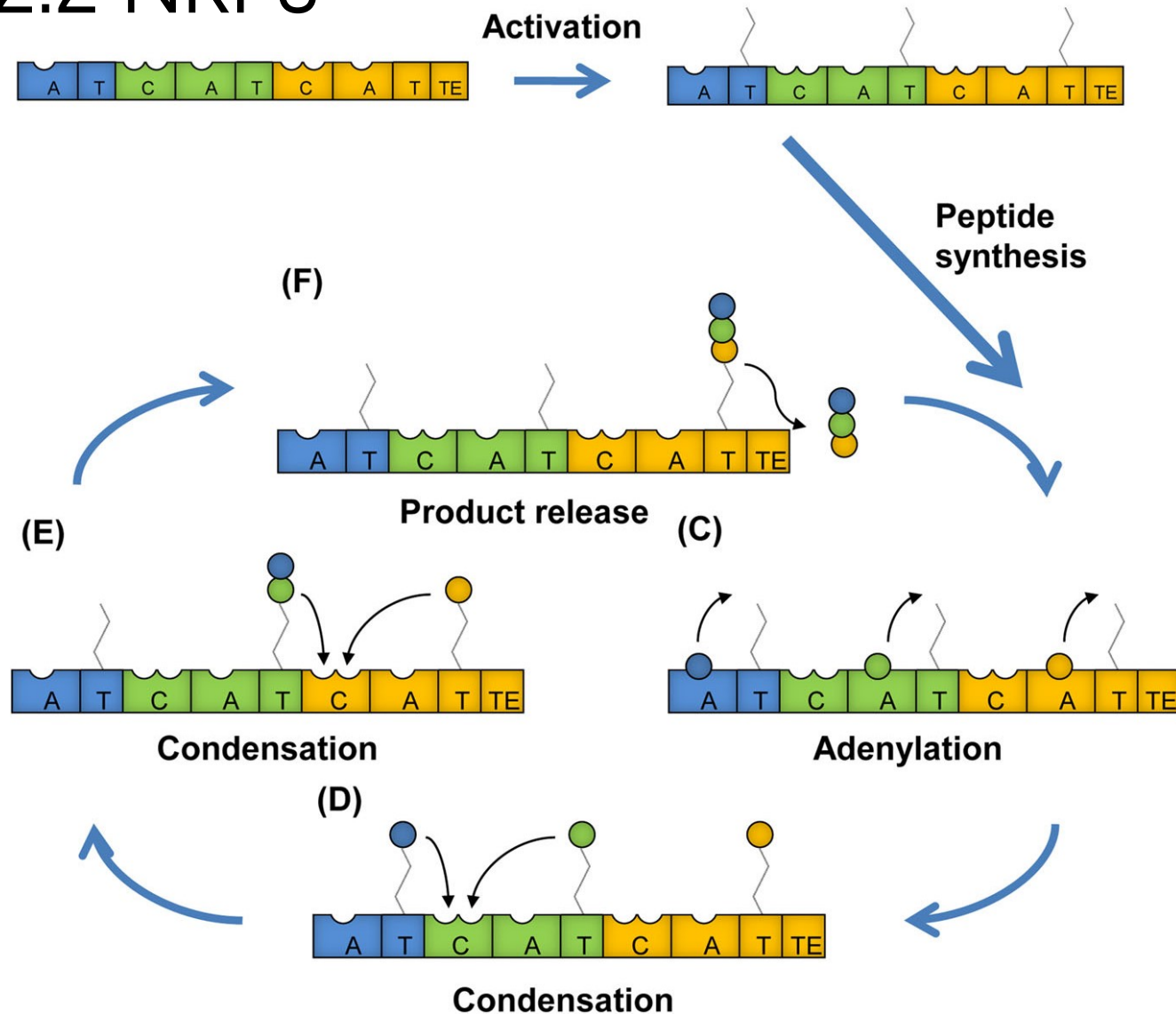
## 2.2 肽类化合物生物合成方式

1. 核糖体合成，经翻译后修饰形成，被称为核糖体合成翻译后修饰多肽 (ribosomal posttranslational peptides, RiPPs)。
2. 利用非核糖体肽合成酶(non-ribosomal peptide synthetase, NRPS)合成分子骨架。NRPS也是一类具有多种功能域的大型合酶。

## 2.2.1 RiPPs

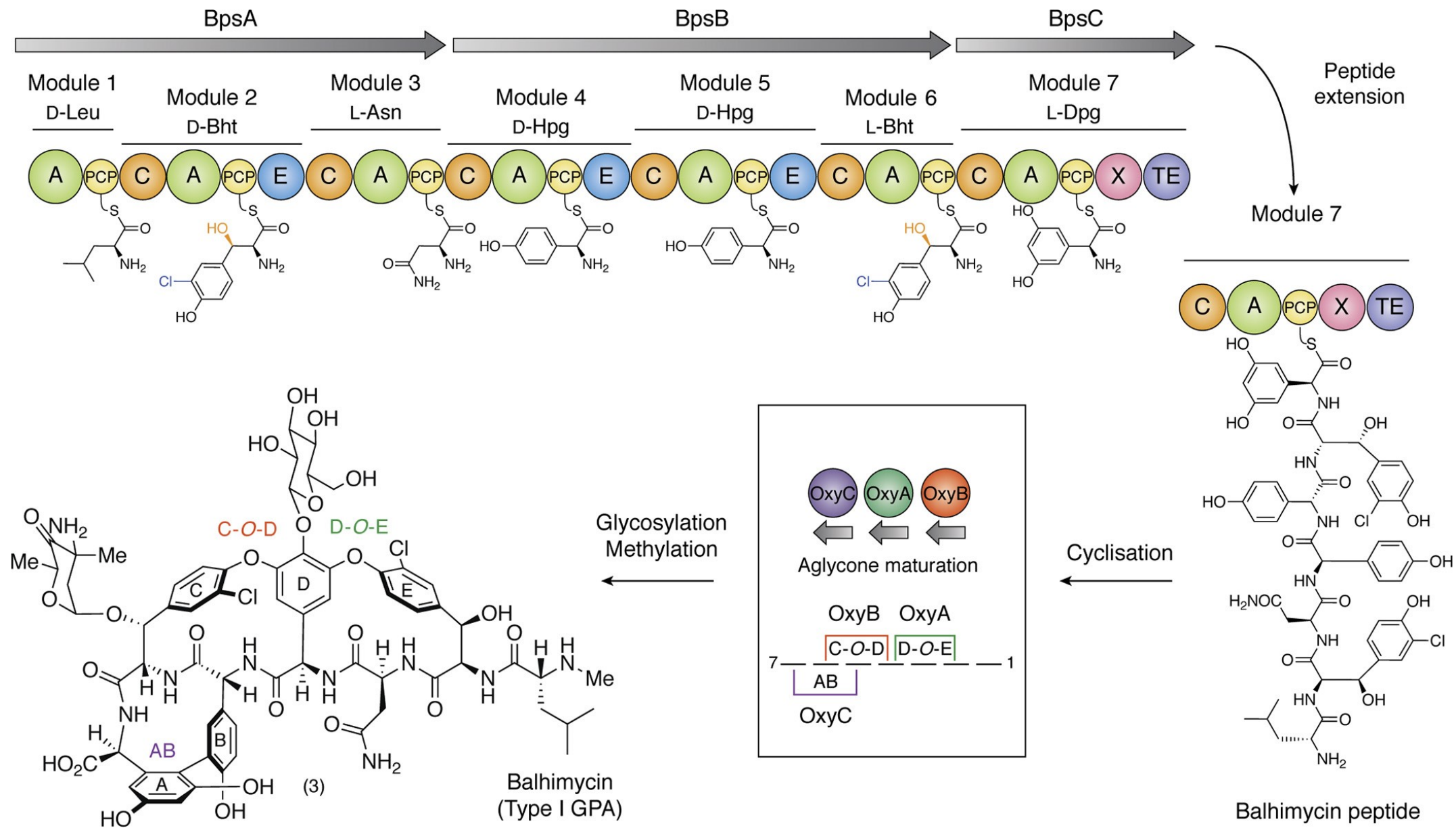


## 2.2.2 NRPS

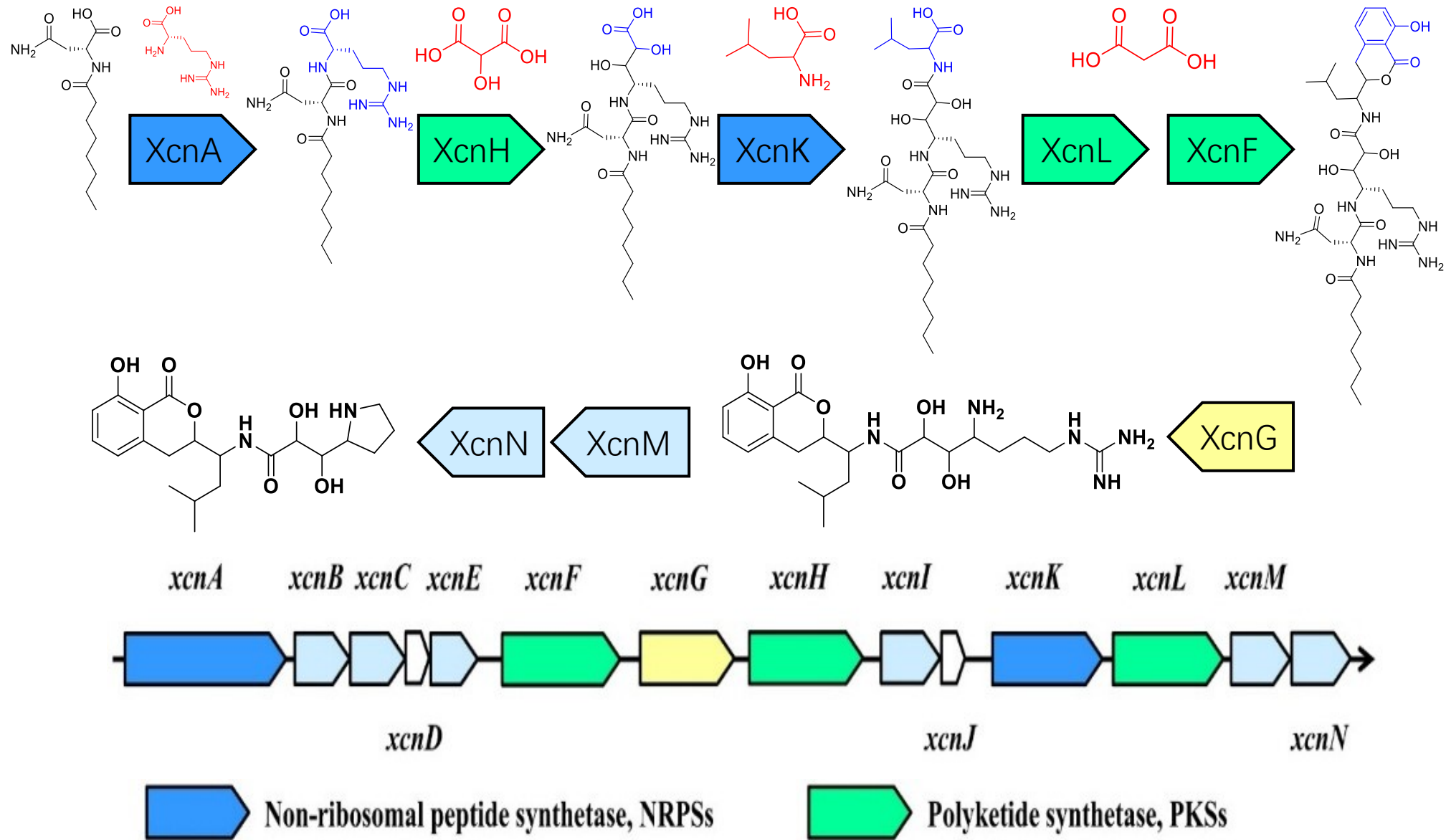


**A:** adenylation domain  
**C:** condensation domain  
**E:** epimerization domain  
**PCP:** peptidyl carrier protein  
**X:** Oxy recruitment domain

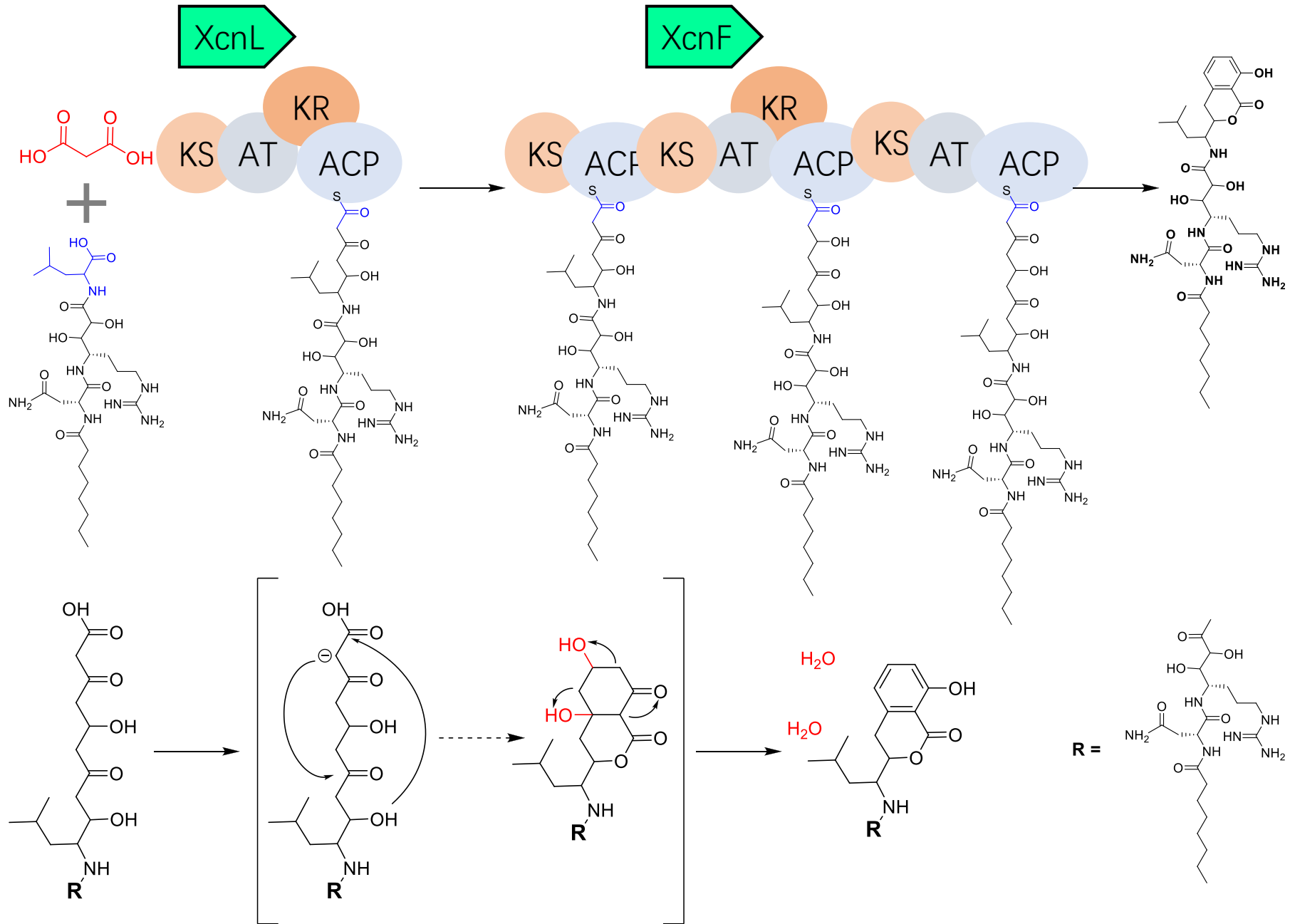
Mark J. Calcott & David F. Ackerley,  
Biotechnol Lett, 2014



## 2.2.3 NRPS-PKS







# 3. Prediction

## 3.1 结构域预测

- <http://smart.embl-heidelberg.de/>
- <https://www.ebi.ac.uk/interpro/>
- <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

## 3.2 基因簇挖掘

- <http://www.secondarymetabolites.org/>
- <https://antismash.secondarymetabolites.org/#!/start>

# 3.1 结构域预测

<http://smart.embl-heidelberg.de/>

**SMART** GENOMES

SMART MODE:  
NORMAL  
GENOMIC

Simple  
Modular  
Architecture  
Research  
Tool

Letunic et al. (2017) *Nucleic Acids Res* doi: 10.1093/nar/gkx922  
Letunic et al. (2020) *Nucleic Acids Res* doi: 10.1093/nar/gkaa937

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

### Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART analysis service.

Sequence ID or ACC

Examples: #1, #2

Protein sequence

Examples: #1, #2

Sequence SMART Reset

HMMER searches of the SMART database occur by default. You may also find:

- Outlier homologues and homologues of known structure
- PFAM domains
- signal peptides
- internal repeats

**SMART** SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

+ = - SAVE

ketoacyl-synt Ketoacyl-synt\_C KAsynt\_C\_assoc PKS\_AT PKS\_KR PP-binding

Architecture

### Domain architecture analysis

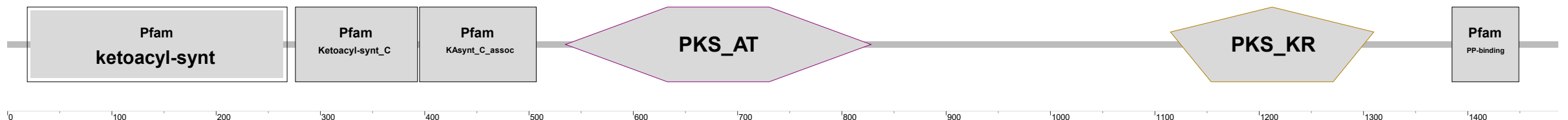
Display all proteins with similar:

- Domain organisation:** Proteins having all the domains as the query in the same order. Additional domains are allowed.
- Domain composition:** Proteins with the same domain composition have at least one copy of each of domains of the query.

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not displayed. The display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > L

**Confidently predicted domains, repeats, motifs and features:**

Name	Start ▲	End	E-value
<a href="#">Pfam:ketoacyl-synt</a>	19	268	1.3e-59
<a href="#">Pfam:Ketoacyl-synt_C</a>	276	393	9e-36
<a href="#">Pfam:KAsynt_C_assoc</a>	395	507	3.2e-19
<a href="#">PKS_AT</a>	535	828	7e-55
<a href="#">PKS_KR</a>	1115	1310	1.57e-25
<a href="#">Pfam:PP-binding</a>	1385	1449	1.1e-8



<https://www.ebi.ac.uk/interpro/>

**InterPro** Classification of protein families

Home Search Browse Results

Search by sequence Search by text

Sequence, in FASTA format

Enter your sequence

Choose file Example protein sequence

Advanced options

Search Clear

**InterPro** Classification of protein families

Home Search Browse Results Release notes Download Help About

None predicted

Entry matches to this protein ⓘ

Options Export

▼ Domain

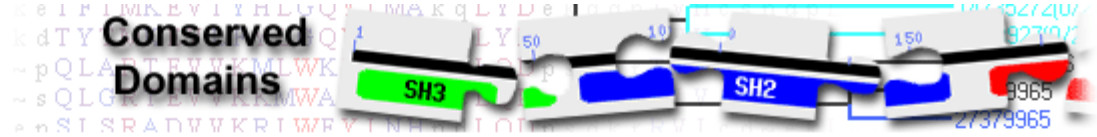
- PP-bd\_ACP
- PP-binding
- CARRIER
- Acyl\_transferas
- PKS\_AT
- Acyl\_transf\_1
- PKS\_assoc
- KAsynt\_C\_assoc
- Ketoacyl\_synth
- ketoacyl-synt
- PKS\_KR
- KR
- PKS\_Beta-keto
- PKS\_KS

Search for

Enter **protein** or **nucleotide** query  
multiple protein queries, use [Batch](#)

Submit

<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

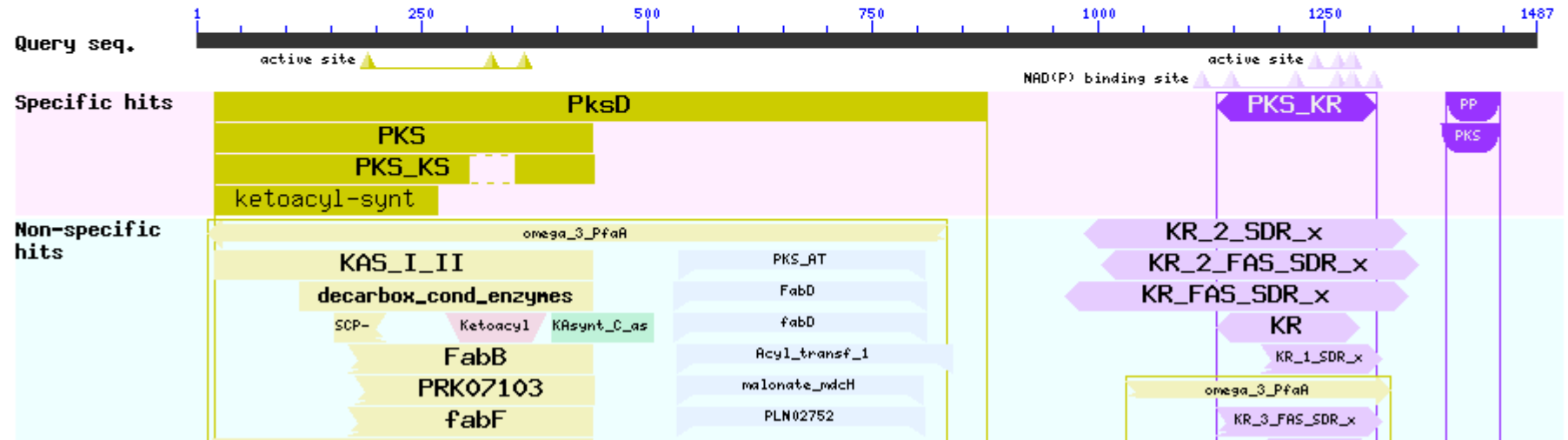


Conserved domains on [lcl|seqsig\_MNKKE\_bda5a18d8fff6f70192d2f27d5d890ed]

View [Full Results](#)

Local query sequence

Graphical summary  Zoom to residue level hide extra options <<  Show site features Horizontal zoom: x 1 Update graph



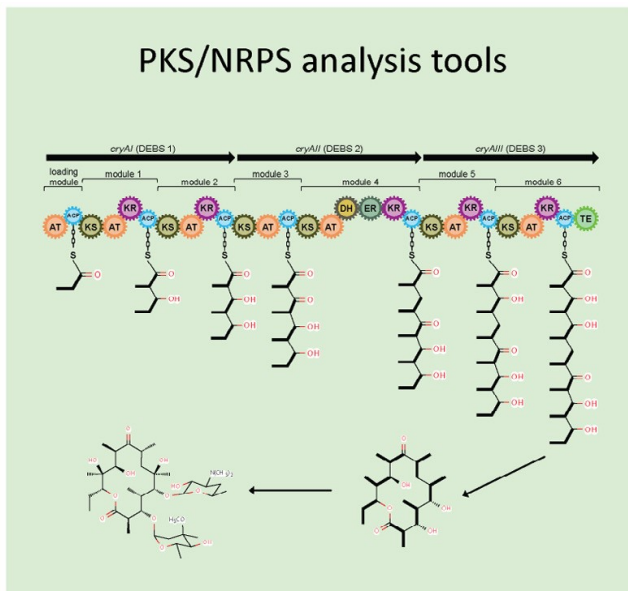
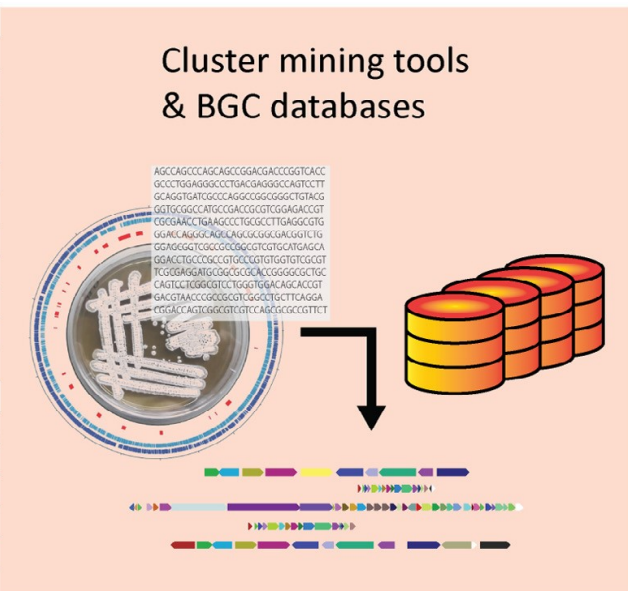
# 3.2 基因簇挖掘

## Tools

- 2metDB
- antiSMASH
- BAGEL
- CLUSEAN
- ClusterFinder
- ClustScan
- eSNaPD
- EvoMining
- GNP/PRISM
- NaPDoS
- SMURF

## Databases

- Bactibase
- ClusterMine360
- CSDB
- DoBISCUIT
- IMG-ABC
- MIBiG

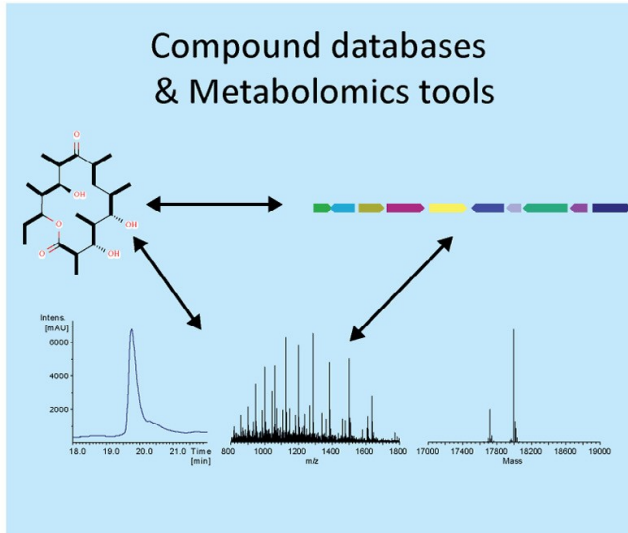
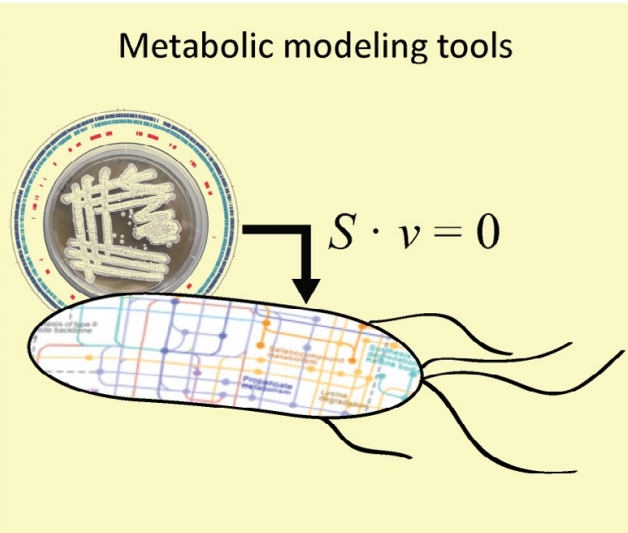


## Tools

- antiSMASH
- ClustScan
- GNP/PRISM
- LSI-based pred.
- MAPSI/ASMPKS
- NaPDoS
- NP.searcher
- NRPS-PKS
- NRPS/PKS SP
- NRPSpredictor
- NRPSsp
- PKS/NRPS WS
- PKSIIIexplorer
- SBSPKS
- SEARCHGT
- SEARCHPKS
- SEQL-NRPS
- SP Type I PKS

## Tools

- antiSMASH
- CoReCo
- FAME
- GEMSiRV
- MEMOSys
- merlin
- MetaFlux
- MicrobesFlux
- Model SEED
- RAVEN
- SuBliMinaL



## Tools

- Cycloquest
- GNPS
- GNP/iSNAP
- NRPquest
- RiPPquest
- Pep2Path

## Databases

- Antibioticome
- ChEBI
- ChEMBL
- ChemSpider
- KNAPSAcK
- NORINE
- Novel Antibiotics
- PubChem
- StreptomeDB

Weber T , Kim H U . Synthetic & Systems Biotechnology, 2016

## Home

The Secondary Metabolite  
Bioinformatics Portal

antiSMASH

antiSMASH database

Contribution

Contributors (in alphabetical  
order)

Citation

antiSMASH

MIBiG

BGC databases

Compounds

Cluster mining tools

Metabolomics tools

Metabolic modeling tools

PKS/NRPS analysis tools

Next »

# Welcome to [www.secondarymetabolites.org](http://www.secondarymetabolites.org/)

## The Secondary Metabolite Bioinformatics Portal

Welcome to the new portal. Here you will find information on all aspects of Secondary Metabolite Bioinformatics, including hand-curated links to all major tools and databases commonly used in the field

### antiSMASH

You are looking for our antiSMASH web-service? Please click [here](#).

Click <http://antismash.secondarymetabolites.org> for a direct link to the service.

### antiSMASH database

You are looking for our antiSMASH-database? Click <http://antismash-db.secondarymetabolites.org> for a direct link to the service.



Server status: **working**

Running jobs: **1**

Queued jobs: **0**

Jobs processed: **833145**

Nucleotide input    **Results for existing job**

Search a genome sequence for secondary metabolite biosynthetic gene clusters

Load sample input

Open example output

### Notification settings

your@email.com

Email address (optional)

### Data input

Upload file    **Get from NCBI**

NCBI acc #

NCBI accession number of desired sequence

### Detection strictness: relaxed



- Detects well-defined clusters containing all required parts.
- Detects partial clusters missing one or more functional parts.

Extra features    All off    All on



Select genomic region:

Overview

1.1

2.1

2.2

2.3

2.4

2.5

2.6

2.7

4.1

4.2

4.3

4.4

5.1

9.1

9.2

10.1

13.1

16.1

18.1

27.1

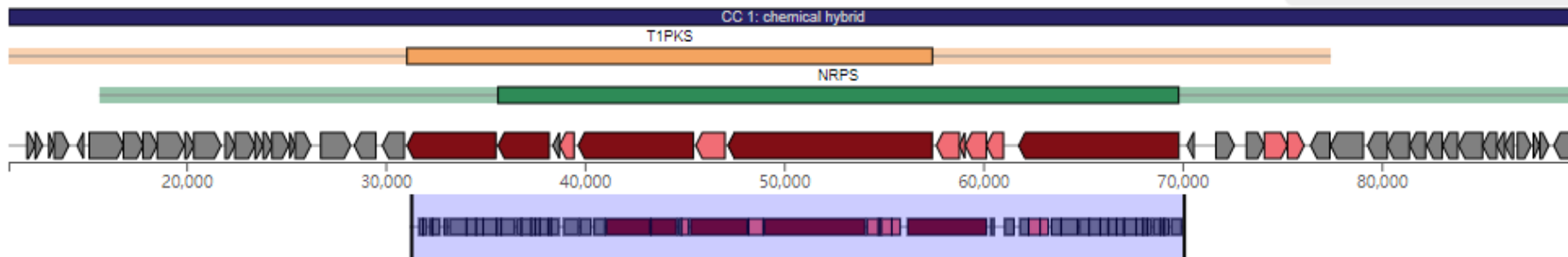
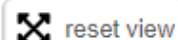
## Identified secondary metabolite regions using strictness 'relaxed'

Compact view



Region	Type	From	To	Most similar known cluster	Similarity
Region 1.1	betalactone <a href="#">↗</a>	569,415	595,271	lipopolysaccharide <a href="#">↗</a>	Saccharide:Lipopolysaccharide 5%
Region 2.1	NRPS <a href="#">↗</a>	1	47,014	xenoamicin A / xenoamicin B <a href="#">↗</a>	NRP:Cyclic depsipeptide 45%
Region 2.2	T1PKS <a href="#">↗</a> , NRPS <a href="#">↗</a>	86,770	144,122		
Region 2.3	NRPS <a href="#">↗</a>	178,506	234,015	nematophin <a href="#">↗</a>	NRP 87%
Region 2.4	NRPS <a href="#">↗</a>	254,851	310,555	nematophin <a href="#">↗</a>	NRP 100%
Region 2.5	bacteriocin <a href="#">↗</a>	312,386	322,589		
Region 2.6	NRPS <a href="#">↗</a>	419,352	491,844	odilorhabdins <a href="#">↗</a>	NRP 100%
Region 2.7	NRPS-like <a href="#">↗</a>	531,623	572,153	xenematide <a href="#">↗</a>	NRP 100%
Region 4.1	NRPS <a href="#">↗</a>	10,787	61,947	xenematide <a href="#">↗</a>	NRP 100%
Region 4.2	NRPS <a href="#">↗</a>	83,686	168,629	xenoamicin A / xenoamicin B <a href="#">↗</a>	NRP:Cyclic depsipeptide 25%
Region 4.3	NRPS-like <a href="#">↗</a>	230,163	271,776		
Region 4.4	NRPS-like <a href="#">↗</a> , thiopeptide <a href="#">↗</a>	315,927	358,986	O-antigen <a href="#">↗</a>	Saccharide 14%
Region 5.1	PpyS-KS <a href="#">↗</a>	223,214	244,260	xenoamicin A / xenoamicin B <a href="#">↗</a>	NRP:Cyclic depsipeptide 8%
Region 9.1	NRPS <a href="#">↗</a>	79,853	126,075	tilivalline <a href="#">↗</a>	NRP 47%
Region 9.2	NRPS-like <a href="#">↗</a>	136,732	178,059	safracin A / safracin B <a href="#">↗</a>	NRP 20%
Region 10.1	T1PKS <a href="#">↗</a> , NRPS <a href="#">↗</a>	11,078	89,813	xenocoumacin 1 / xenocoumacin II <a href="#">↗</a>	NRP + Polyketide:Modular type I 100%
Region 13.1	nucleoside <a href="#">↗</a>	62,440	84,566	TP-1161 <a href="#">↗</a>	RiPP:Thiopeptide 20%
Region 16.1	siderophore <a href="#">↗</a>	2,311	26,843	putrebactin / avaroferrin <a href="#">↗</a>	Other 100%
Region 18.1	NRPS <a href="#">↗</a>	1	44,272	xenoamicin A / xenoamicin B <a href="#">↗</a>	NRP:Cyclic depsipeptide 33%

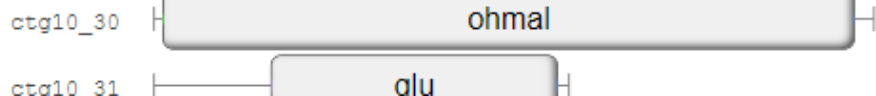
Select genomic region:

[Overview](#)[1.1](#)[2.1](#)[2.2](#)[2.3](#)[2.4](#)[2.5](#)[2.6](#)[2.7](#)[4.1](#)[4.2](#)[4.3](#)[4.4](#)[5.1](#)[9.1](#)[9.2](#)[10.1](#)[13.1](#)[16.1](#)[18.1](#)[27.1](#)**NZ\_HF952051.1 - Region 1 - NRPS, T1PKS****Gene details**Location: 11,078 - 89,813 nt. (total: 78,736 nt) [Show pHMM detection rules used](#)[Download region GenBank file](#)**Legend:** core biosynthetic genes additional biosynthetic genes transport-related genes regulatory genes other genes resistance

reset view



zoom to selection

[NRPS/PKS domains](#)[ClusterBlast](#)[KnownClusterBlast](#)[SubClusterBlast](#)**Detailed domain annotation**Selected features only Show module domains 

ctg10\_30

Locus tag: ctg10\_30

Protein ID: None

Gene: None

Location: 31,078 - 35,541, (total: 4464 nt)

biosynthetic (rule-based-clusters) T1PKS:

PKS\_AT

biosynthetic (rule-based-clusters) T1PKS:

PKS\_KS

biosynthetic-additional (rule-based-clusters) PP-binding

biosynthetic-additional (smcogs)

SMCOG1093:Beta-ketoacyl synthase (Score: 68.1; E-value: 1e-20)

Active site details:

[NCBI BlastP on this gene](#)[View genomic context](#)[MIBIG Hits](#)AA sequence: [Copy to clipboard](#)Nucleotide sequence: [Copy to clipboard](#)[NRPS/PKS products](#)[NRPS/PKS monomers](#)**Predicted core structure(s)**

For candidate cluster 1, location 11077 - 89813:

## 4. Outlook

通过将生物信息学与现代生物技术相结合，加快了SMs及其BGCs研究进程，加深科研人员对生物代谢行为的理解。当前对于SMs仍处于活性成分挖掘，分子基团的简单修饰改造阶段，相信随着研究的深入，我们能够从头设计全新的BGCs，生产新型化学结构的SMs。

# 参考文献



1. Weissman K J . Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology[J]. Natural Product Reports, 2016, 33(2):203-230.
2. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature.[J]. Natural Product Reports, 2013.
3. Calcott M J , Ackerley D F . Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products[J]. Biotechnology Letters, 2014, 36(12):2407-2416.
4. Marschall E , Cryle M J , Tailhades J . Biological, chemical, and biochemical strategies for modifying glycopeptide antibiotics[J]. Journal of Biological Chemistry, 2019, 294(49):jbc.REV119.006349.
5. Reimer D , Pos K M , Thines M , et al. A natural prodrug activation mechanism in nonribosomal peptide synthesis.[J]. Nature Chemical Biology, 2011, 7(12):888-890.
6. Weber T , Kim H U . The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production[J]. Synthetic & Systems Biotechnology, 2016, 1( 2):69-79.
7. Kai, Blin, Simon, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.[J]. Nucleic acids research, 2019.
8. Kai B , Thomas W , Chevrette M G , et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification[J]. Nuclc Acids Research, 2017, 45(Web Server issue).
9. Tilmann W , Kai B , Srikanth D , et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters[J]. Nuclc Acids Research, 2015, 43(W1):W237-W243.

感谢批评指正