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

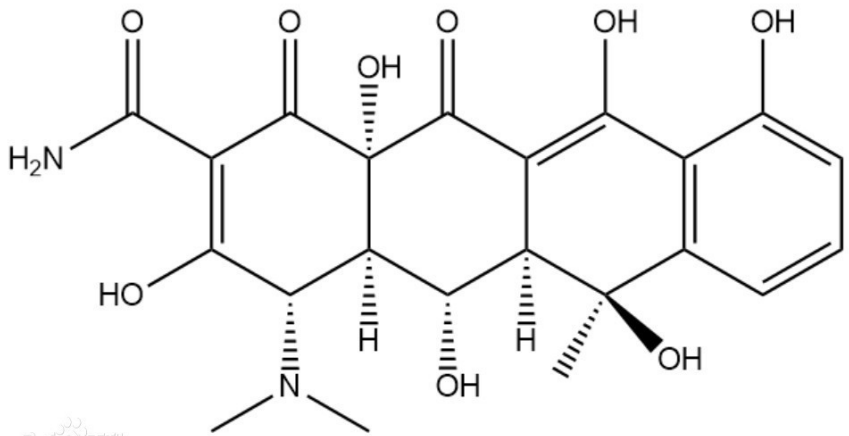
黄花蒿

人参

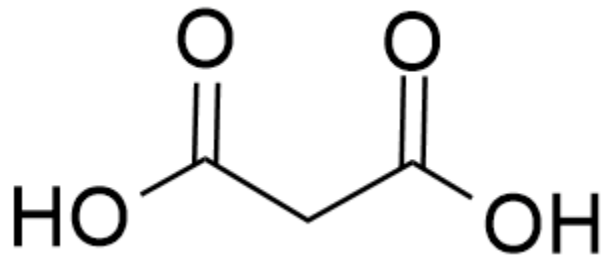
图片来源：百度百科

1.1 聚酮类化合物 (Polyketones,PKs)

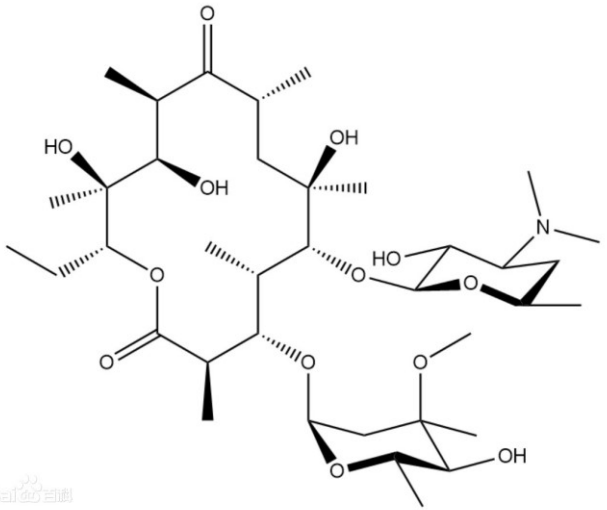
- 土霉素



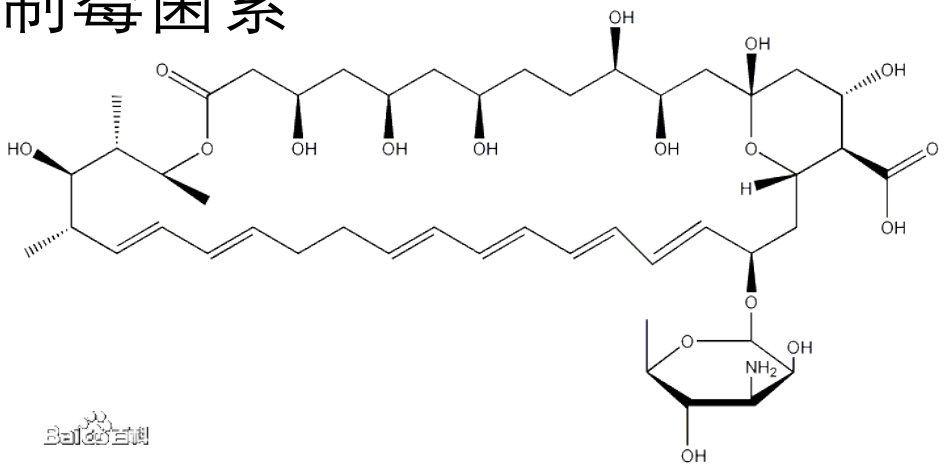
- 基本合成单元



- 红霉素



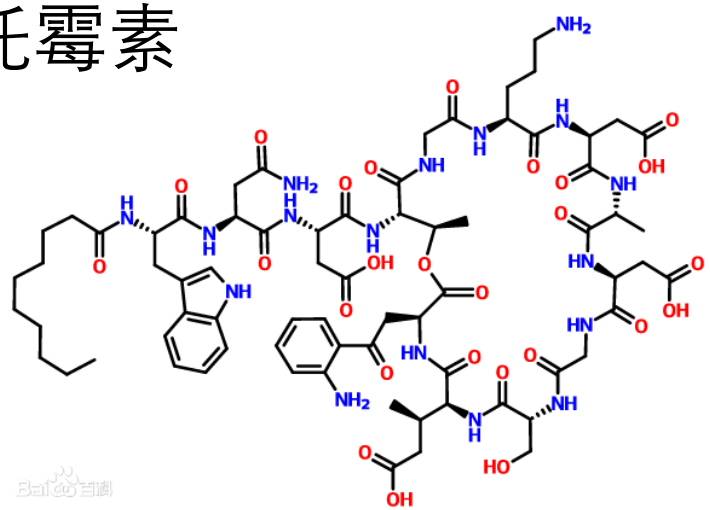
- 制霉菌素



图片来源：百度百科

1.2 肽类化合物 (Petides)

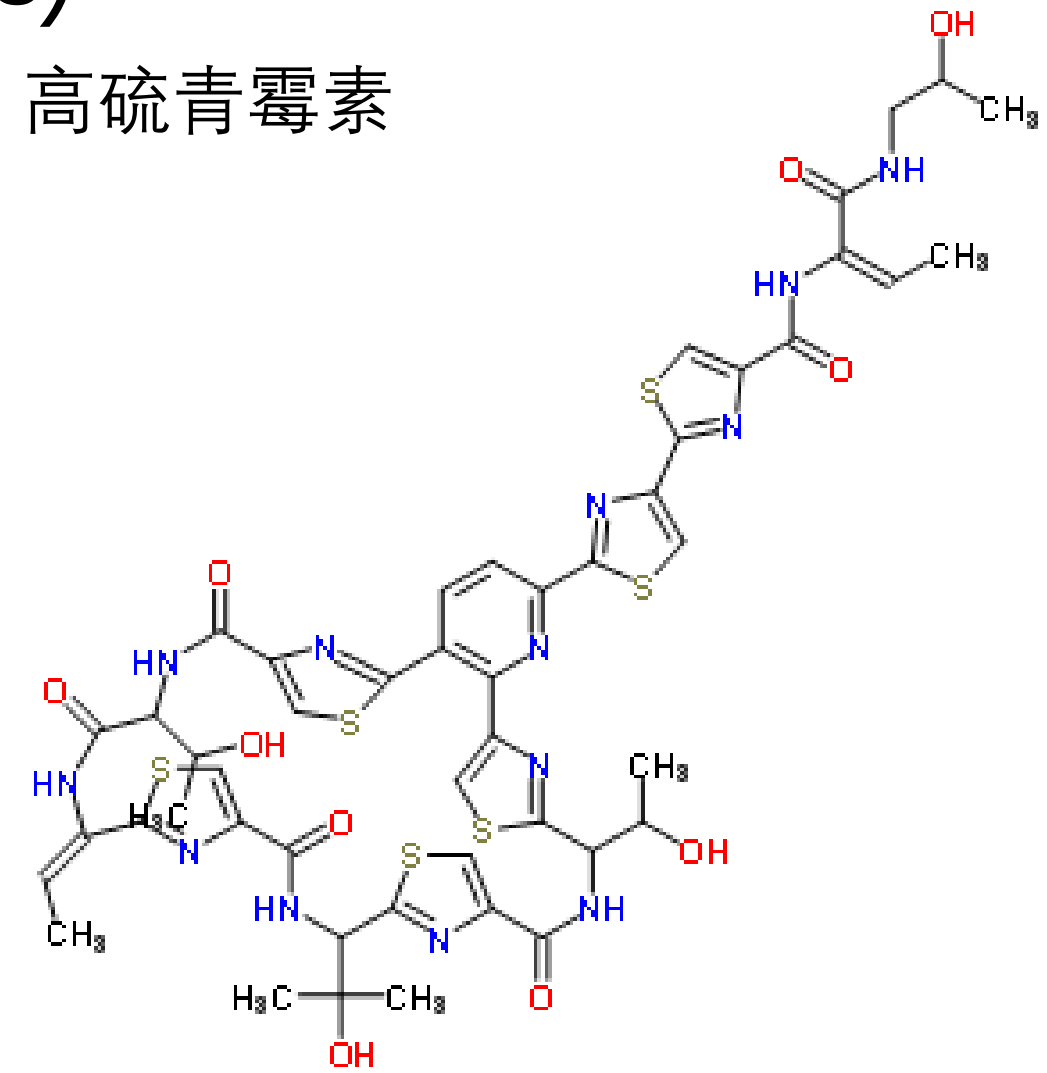
- 达托霉素



- 基本合成单元

天然氨基酸与非天然氨基酸

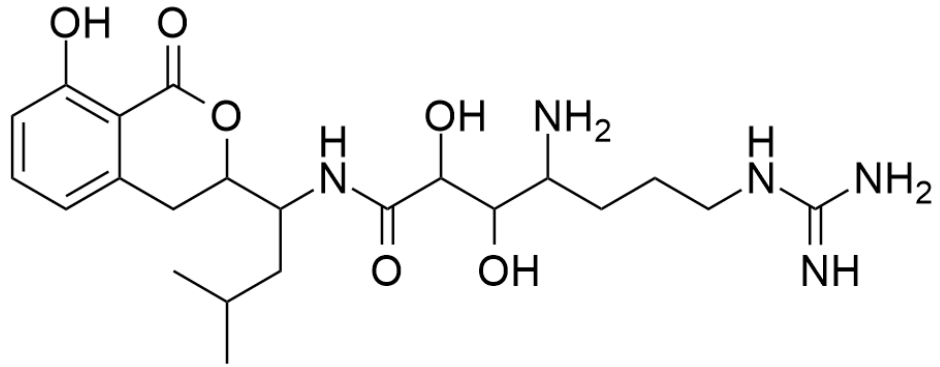
- 高硫青霉素



图片来源：百度百科

1.3 NRP-PK杂合化合物

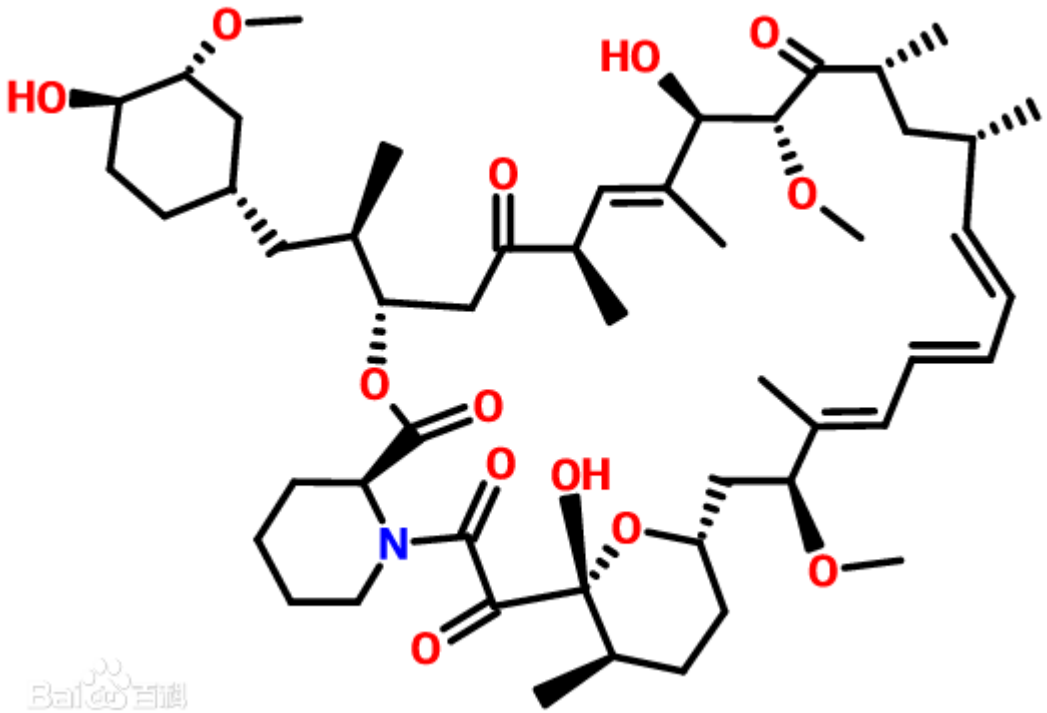
- Xenocoumacin, Xcn1



- 基本合成单元

丙二酸、氨基酸等

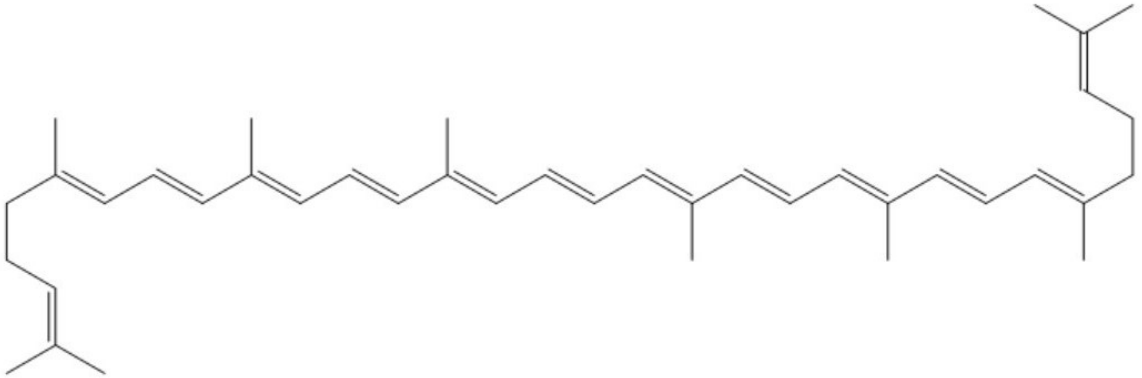
- 雷帕霉素



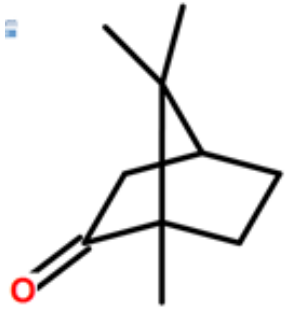
图片来源：百度百科

1.4 异戊二烯类/萜类化合物

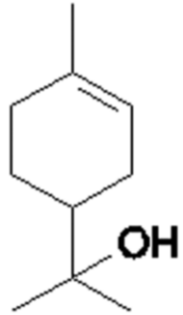
- 番茄红素



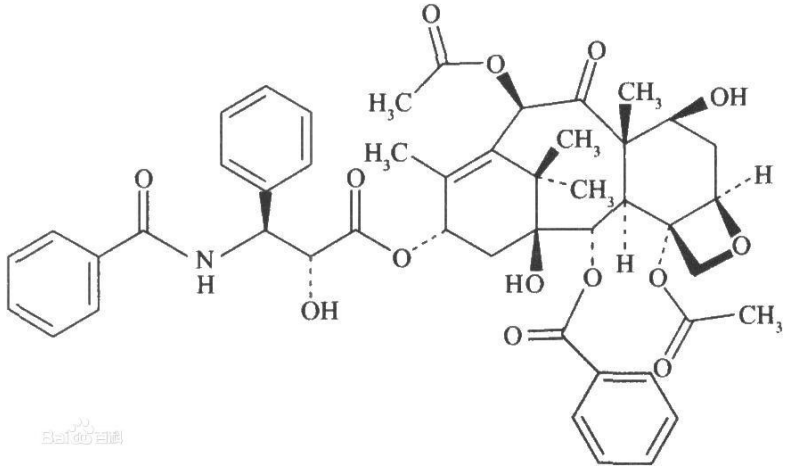
樟脑



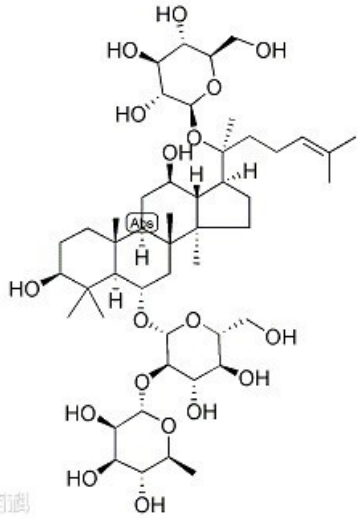
松油醇



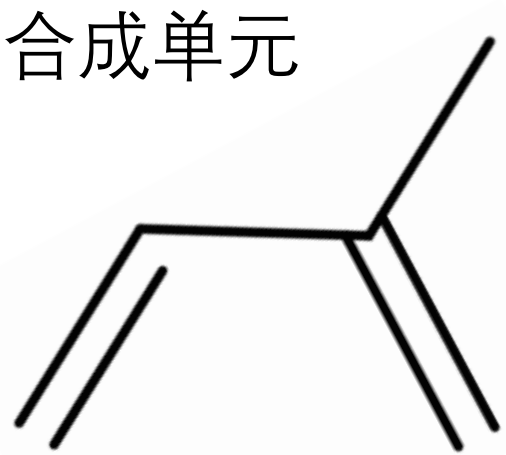
紫杉醇



人参皂苷



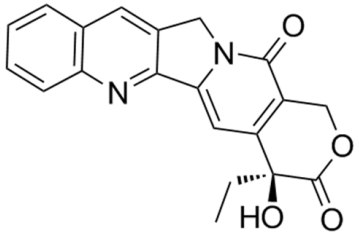
- 基本合成单元



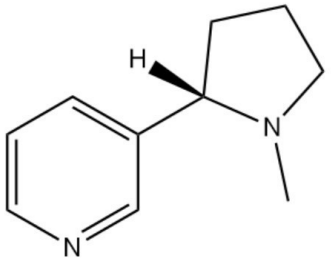
图片来源：百度百科

1.5 生物碱及其它化合物

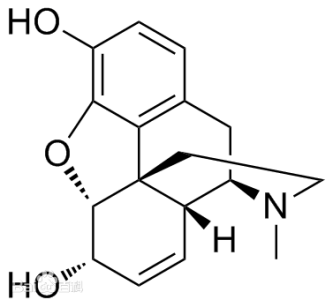
喜树碱



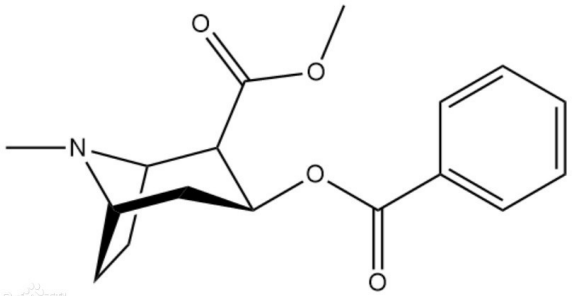
尼古丁



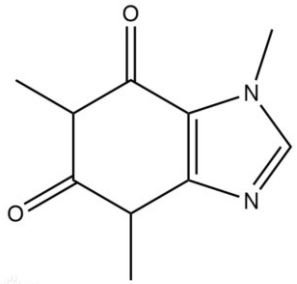
吗啡



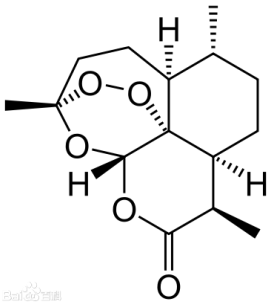
可卡因



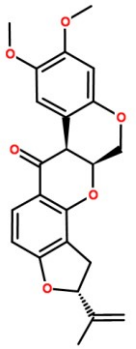
咖啡因



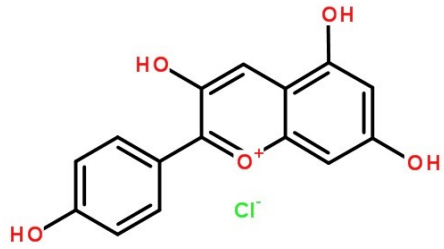
青蒿素



鱼藤酮



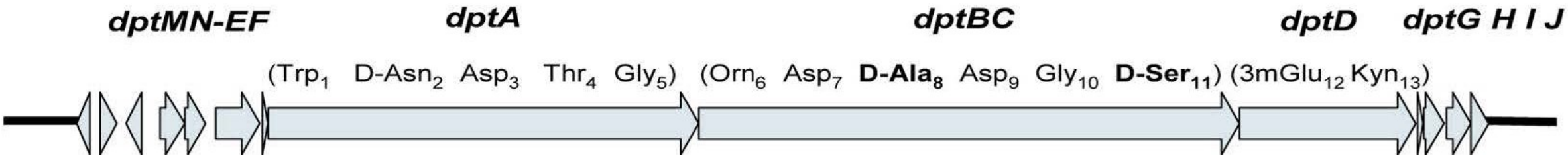
花青素



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 carder protein

TE: thioesterase

KS: keto-acyl synthase

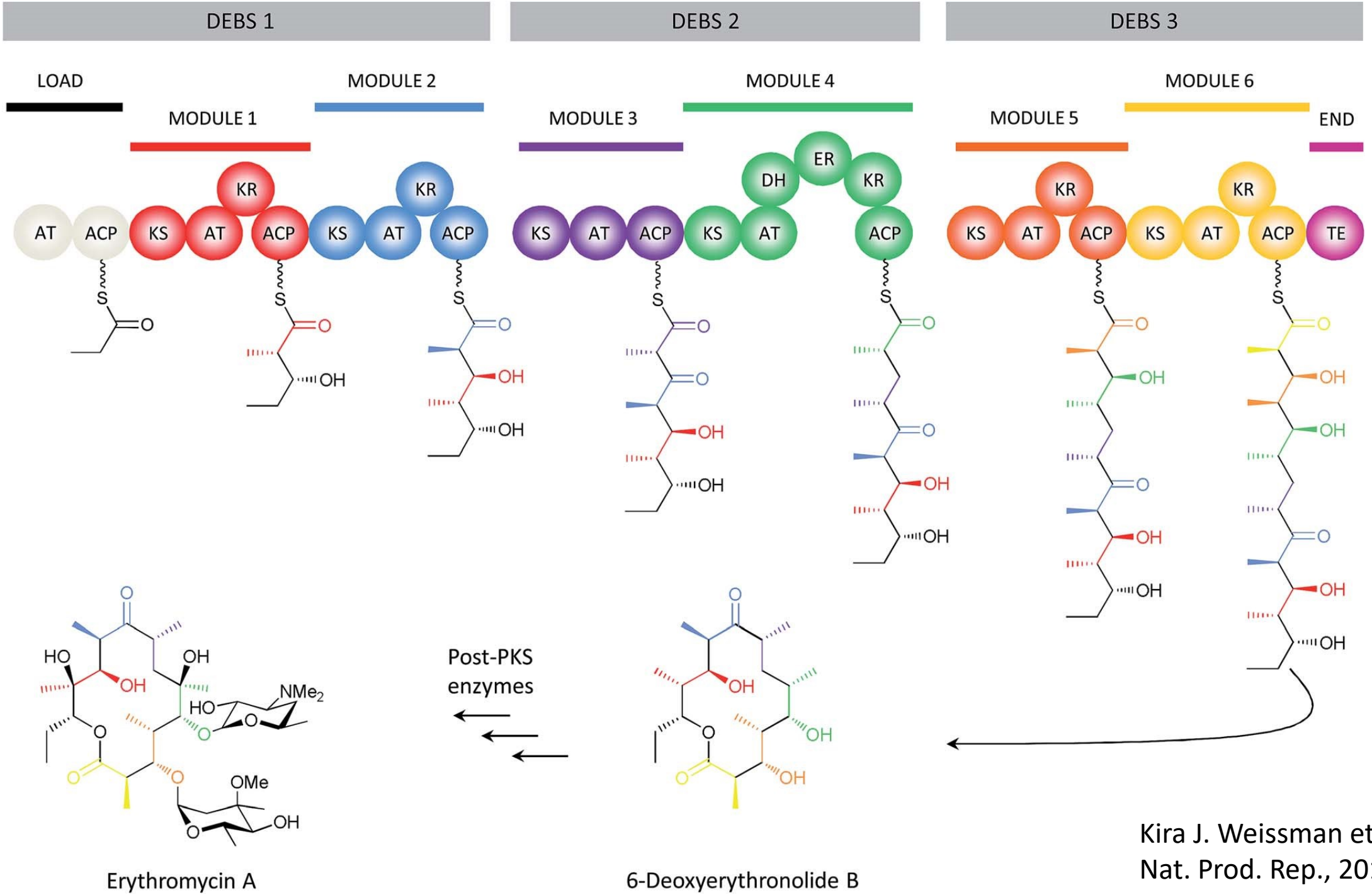
KR: ketoacyl reductase

DH: dehydratase

ER: enoylreductase

MT: meyl transferase

Cy: cyclase



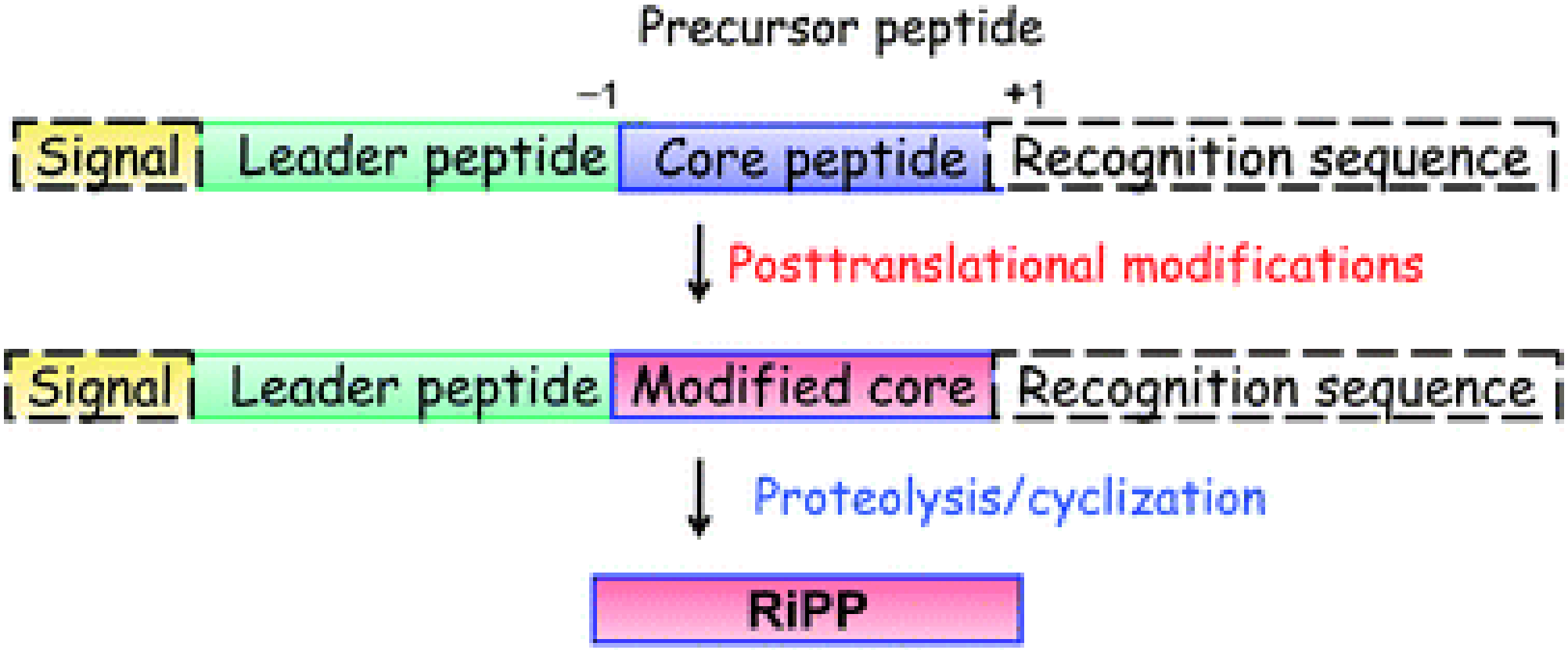
Kira J. Weissman et al.,
Nat. Prod. Rep., 2016

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)。

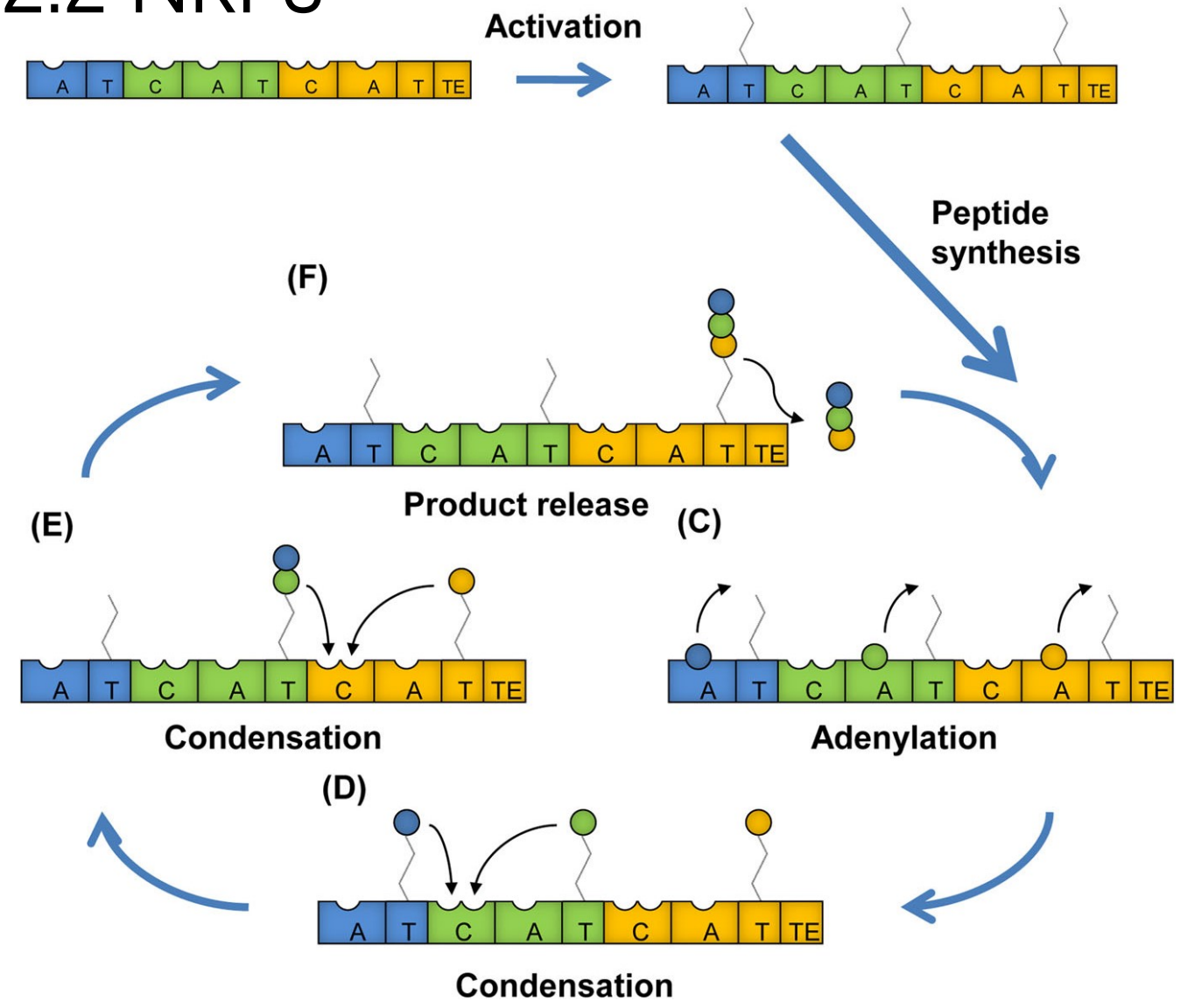
2. 利用非核糖体肽合成酶(non-ribosomal peptide synthetase，NRPS)合成分子骨架。NRPS也是一类具有多种功能域的大型合酶。

2.2.1 RiPPs



Paul G. Arnison et al, Nat. Prod. Rep., 2013

2.2.2 NRPS



A: adenylation domain

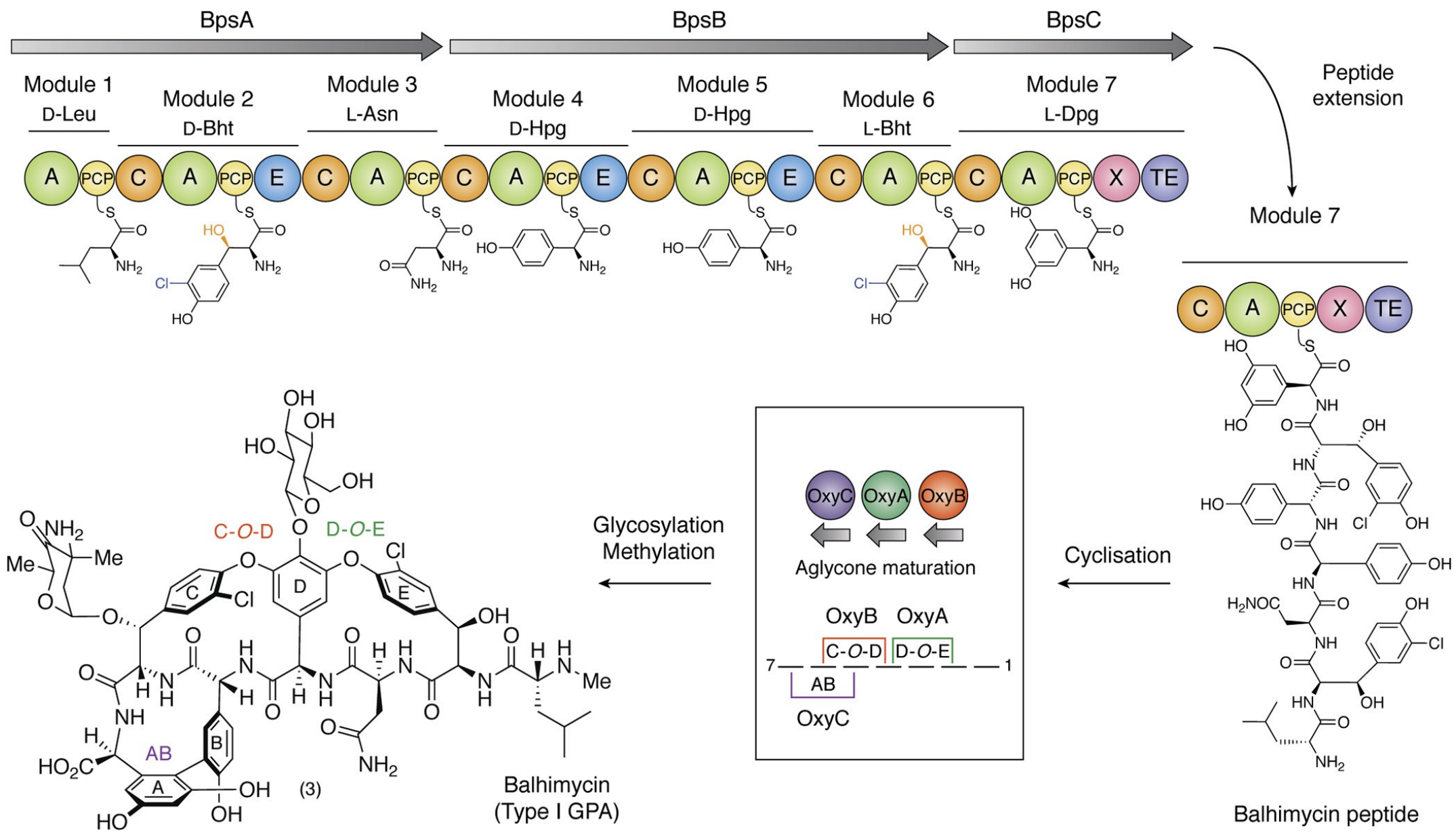
C: condensation domain

E: epimerization domain

PCP: peptidyl carrier protein

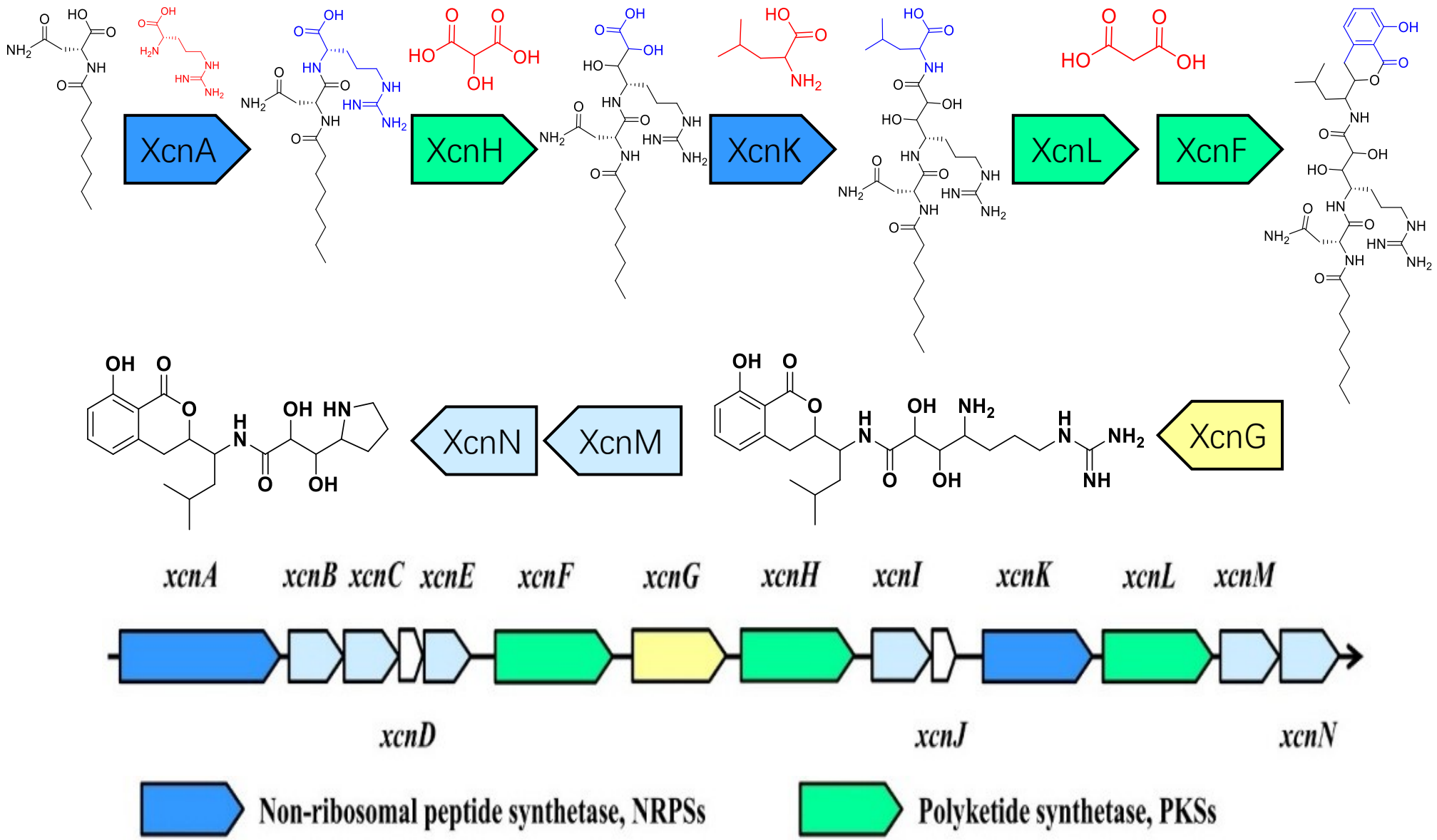
X: Oxy recruitment domain

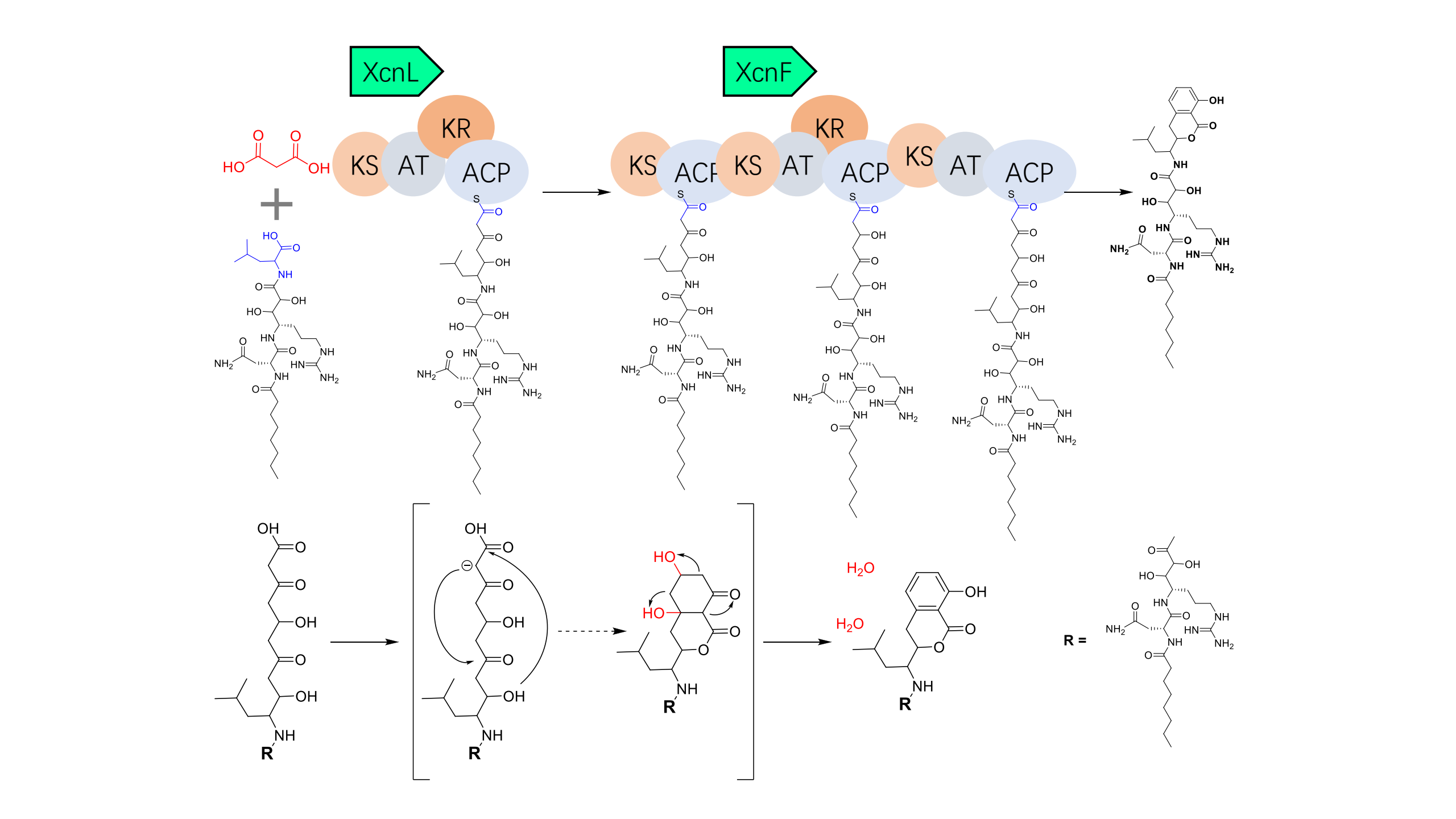
Mark J. Calcott & David F. Ackerley, Biotechnol Lett, 2014



Edward Marschall et al., J. Biol. Chem., 2019

2.2.3 NRPS-PKS


XcnA
XcnH
XcnK
XcnL
XcnF
XcnN
XcnM
XcnG
xcnA
xcnB
xcnC
xcnE
xcnF
xcnG
xcnH
xcnI
xcnK
xcnL
xcnM
xcnD
xcnJ
xcnN
Non-ribosomal peptide synthetase, NRPSs
Polyketide synthetase, PKSs


XcnL
KS
AT
ACP
KR
XcnF
KS
ACP
KS
AT
ACP
KR
KS
AT
ACP
R =

3. Prediction

3.1 结构域预测

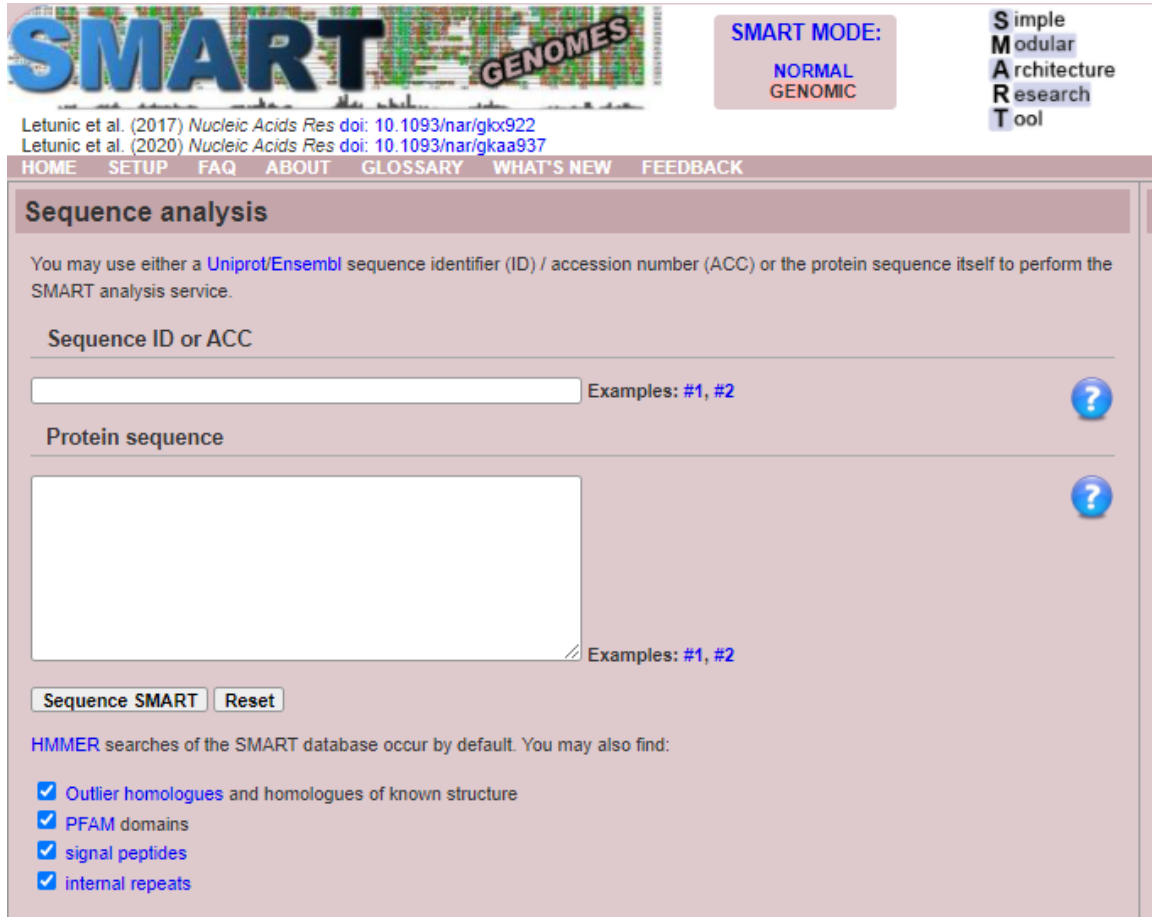
- http://smart.embl-heidelberg.de/
- https://www.ebi.ac.uk/interpro/
- https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

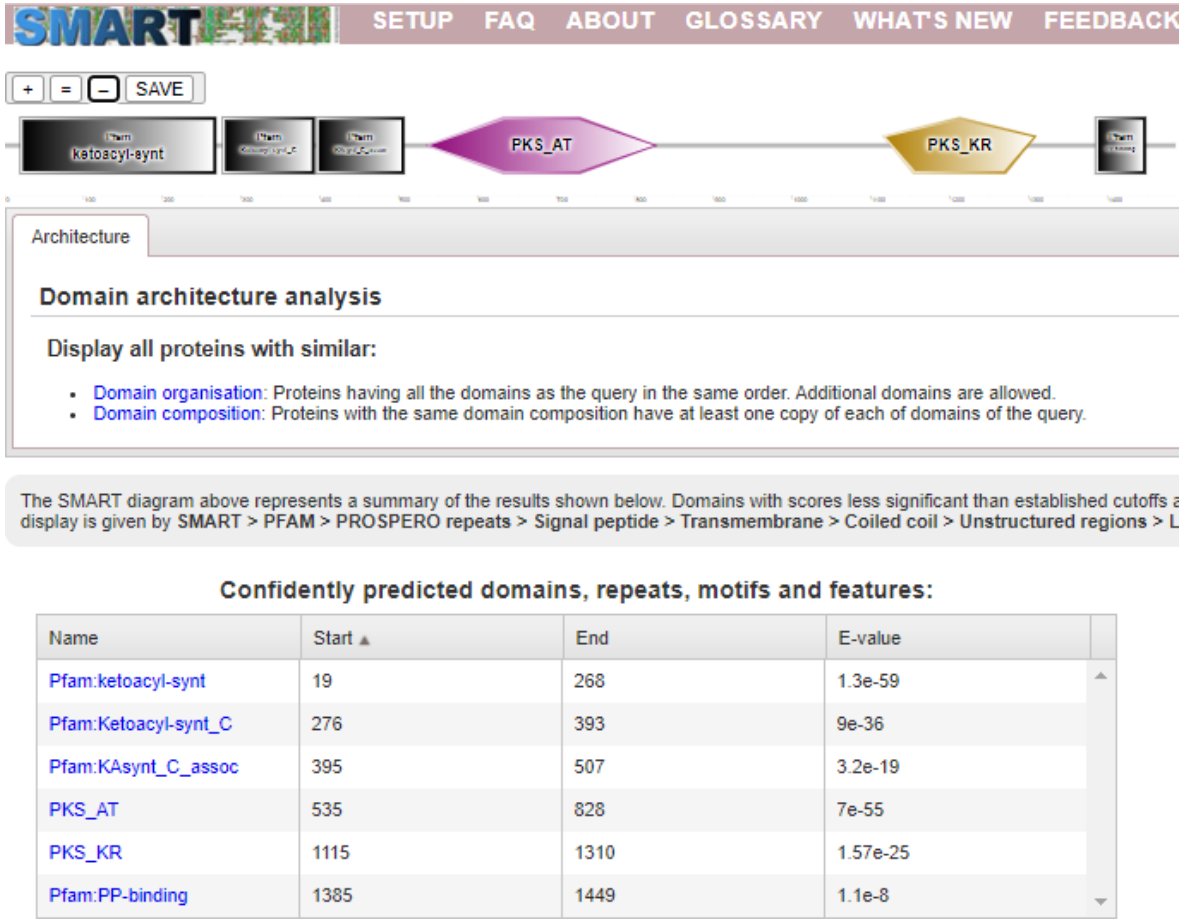
3.2 基因簇挖掘

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

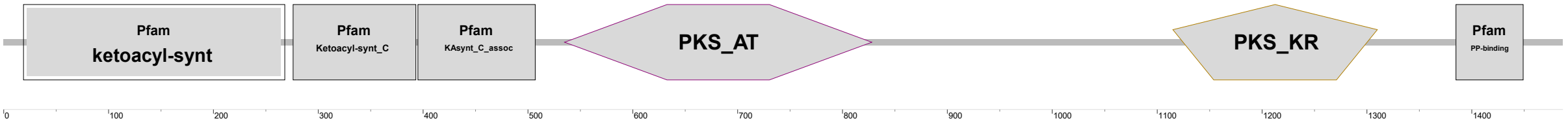
3.1 结构域预测

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

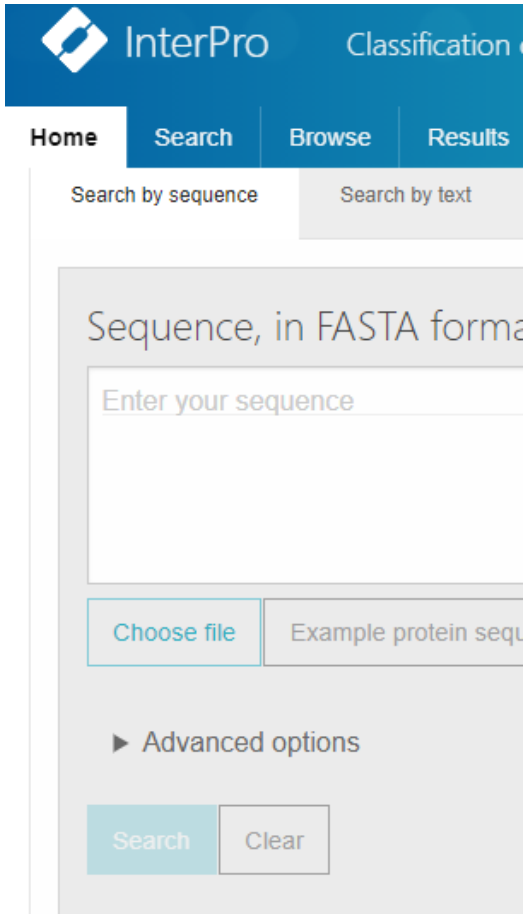


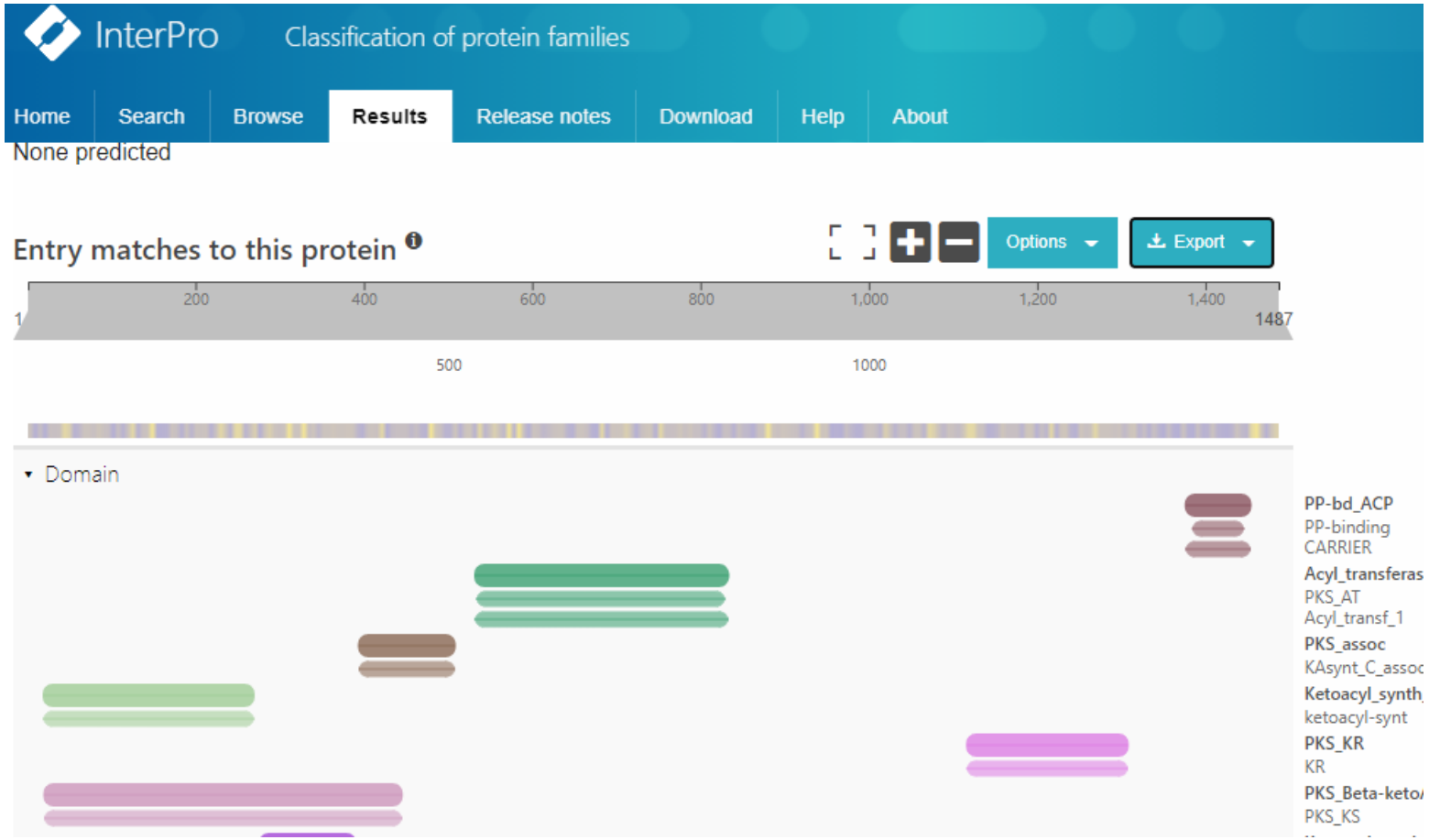


Name	Start ▲	End	E-value
Pfam:ketoacyl-synt	19	268	1.3e-59
Pfam:Ketoacyl-synt_C	276	393	9e-36
Pfam:KAsynt_C_assoc	395	507	3.2e-19
PKS_AT	535	828	7e-55
PKS_KR	1115	1310	1.57e-25
Pfam:PP-binding	1385	1449	1.1e-8

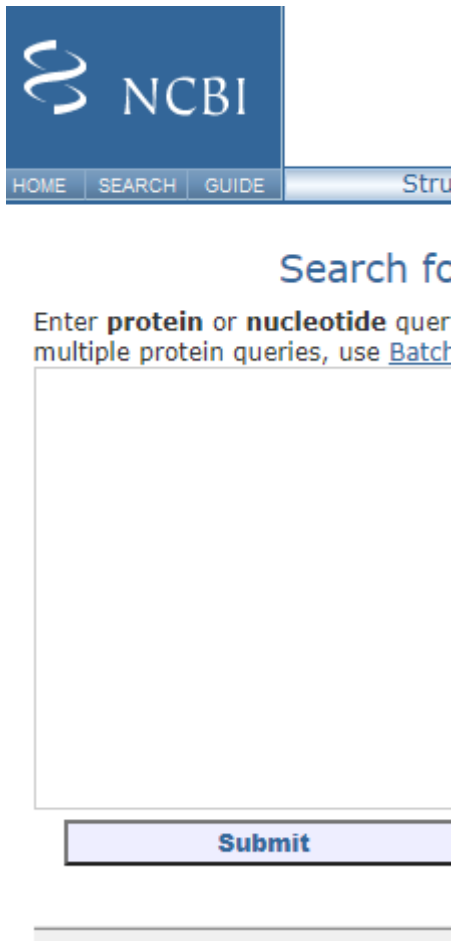


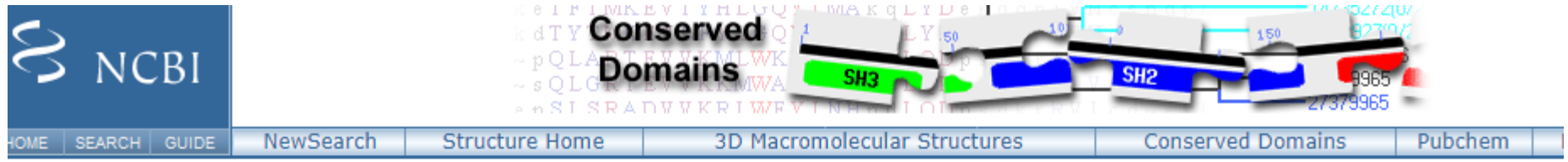
https://www.ebi.ac.uk/interpro/

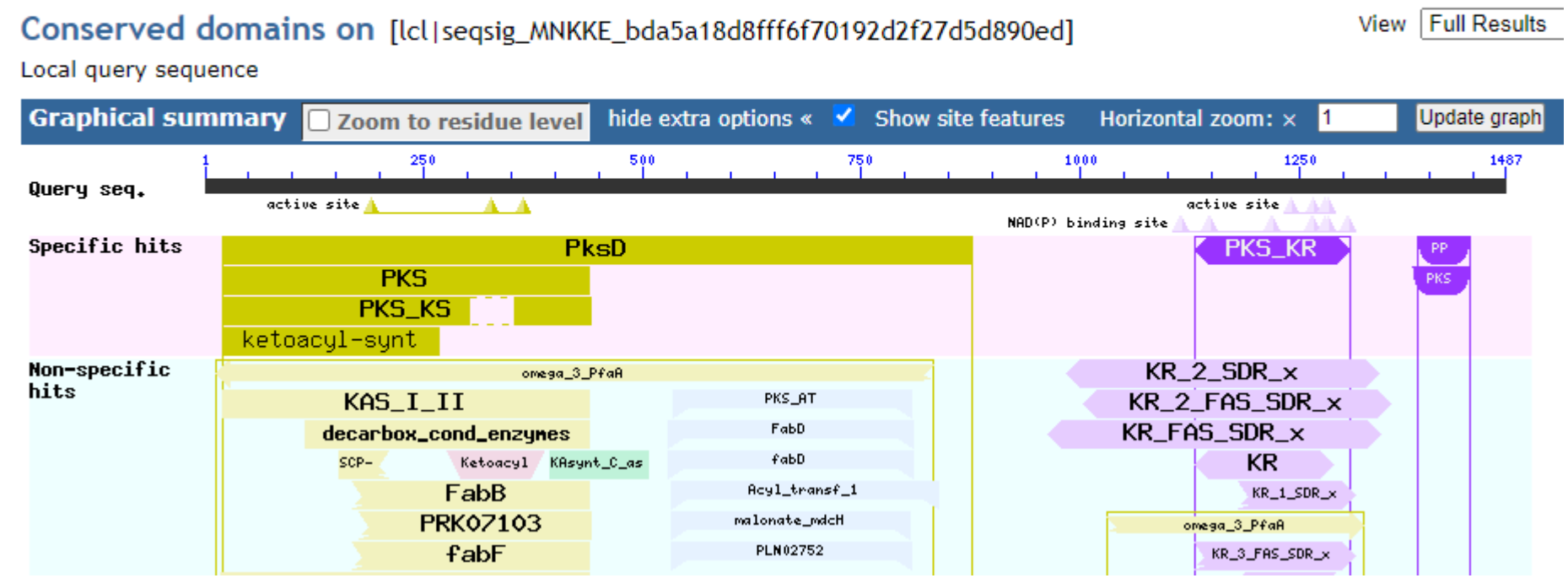




https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi



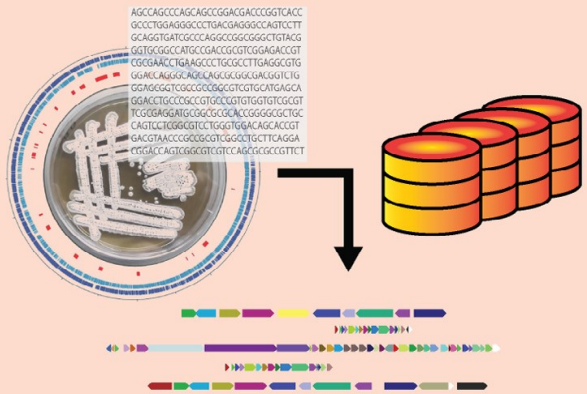




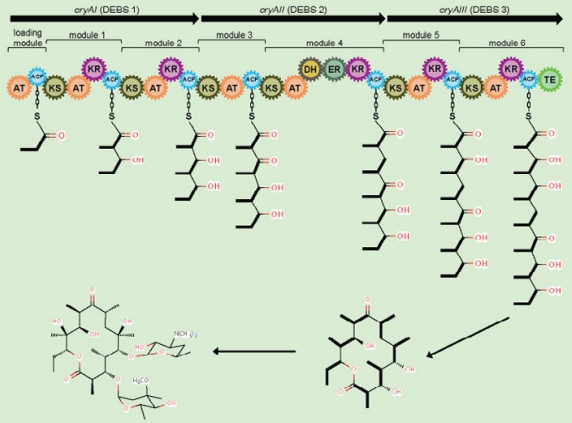
3.2 基因簇挖掘

Tools
2metDB
antiSMASH
BAGEL
CLUSEAN
ClusterFinder
ClustScan
eSNaPD
EvoMining
GNP/PRISM
NaPDoS
SMURF
Databases
Bactibase
ClusterMine360
CSDB
DoBISCUIT
IMG-ABC
MIBiG

Cluster mining tools & BGC databases



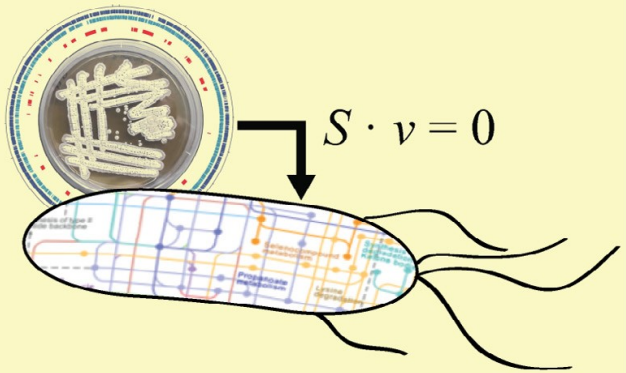
PKS/NRPS analysis tools



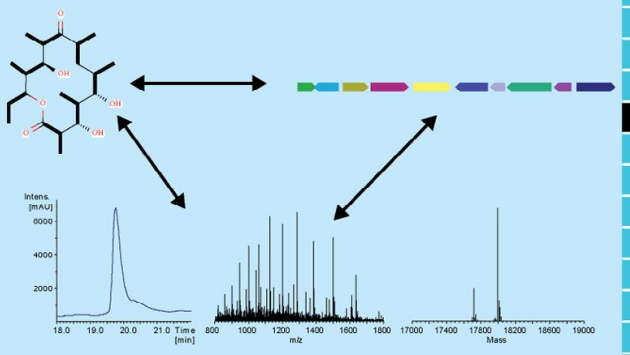
Tools
antiSMASH
ClustScan
GNP/PRISM
LSI-based pred.
MAPSI/ASMPKS
NaPDoS
NP.searcher
NRPS-PKS
NRPS/PKS SP
NRPSpredictor
NRPSsp
PKS/NRPS WS
PKSIIIexplorer
SBSPKS
SEARCHGTr
SEARCHPKS
SEQL-NRPS
SP Type I PKS

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

Metabolic modeling tools



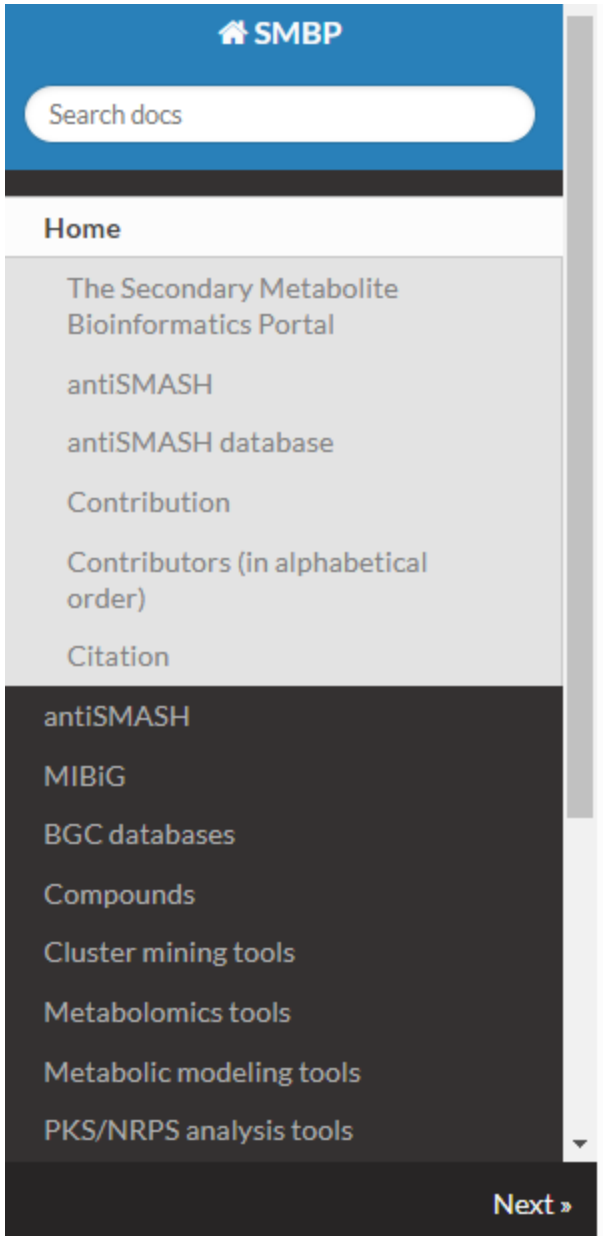
Compound databases & Metabolomics tools



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

http://www.secondarymetabolites.org/



Docs » Home

Welcome to 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 linkts to all major tools and databases commonly used in the field

antiSMASH

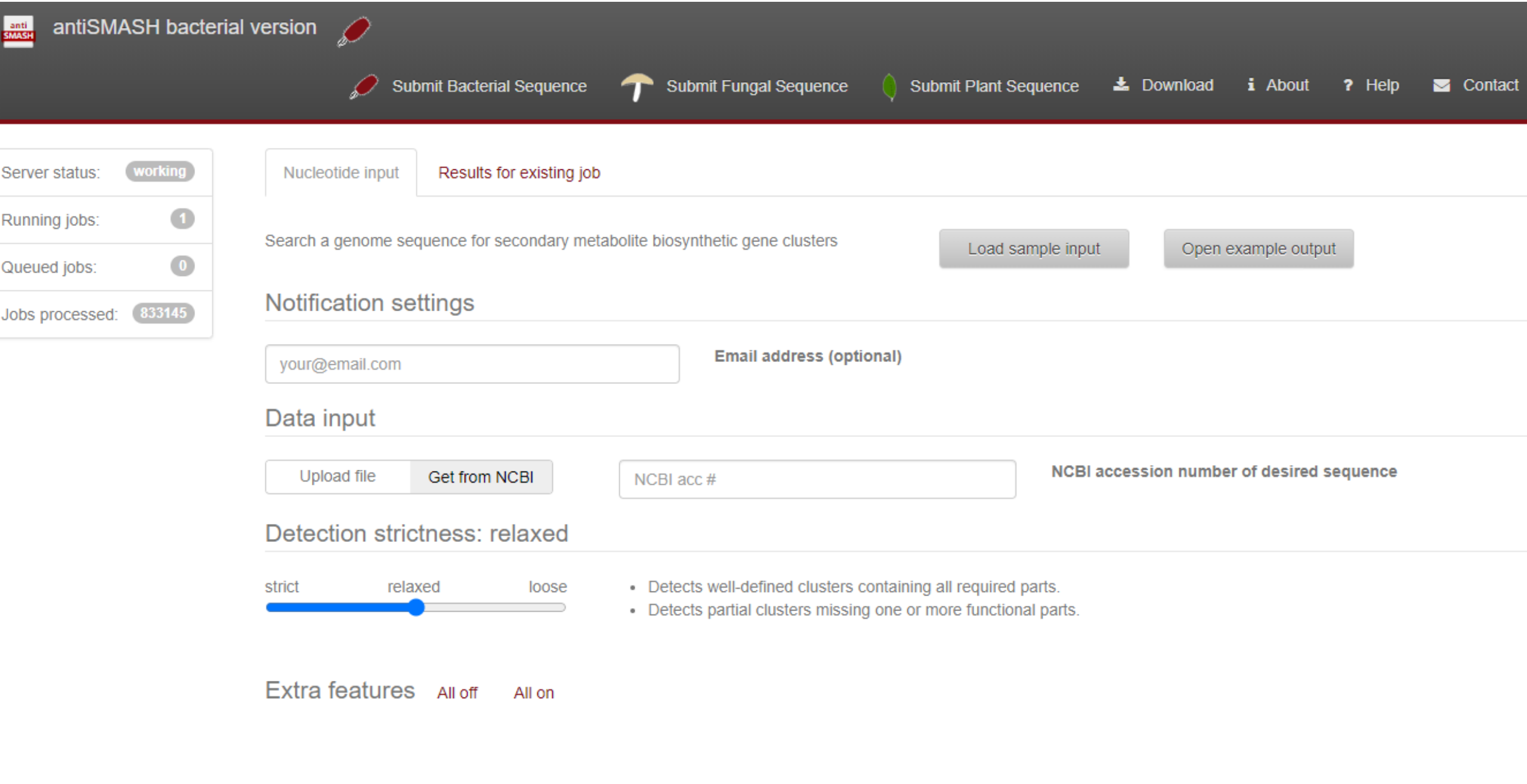
You are looking for our antiSMASH web-service? Please click here.

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.

https://antismash.secondarymetabolites.org/#!/start



antiSMASH version 5.2.0 | Download | About | Help | Contact

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

antiSMASH version 5.2.0 | Download | About | Help | Contact

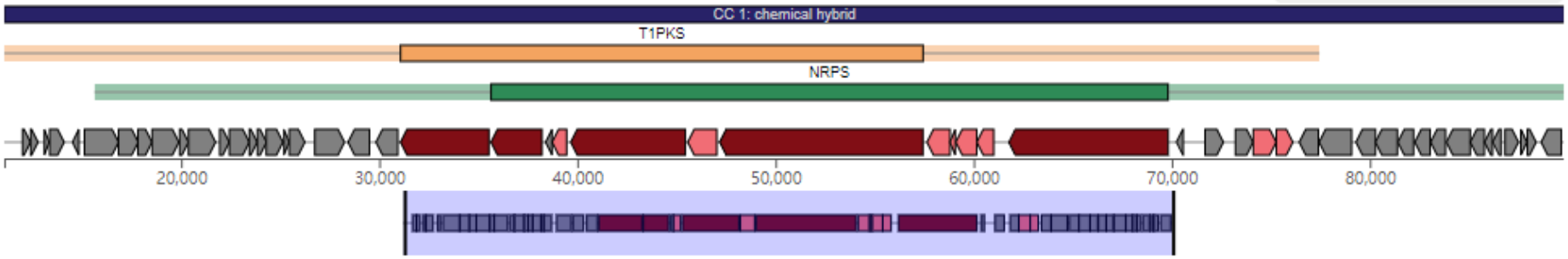
Select genomic region:

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

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

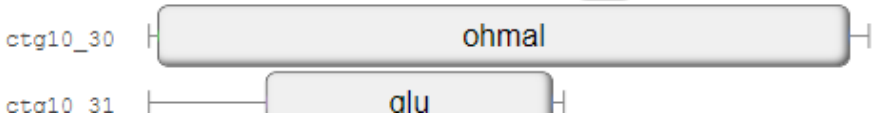
reset view | zoom to selection

NRPS/PKS domains | ClusterBlast | KnownClusterBlast | SubClusterBlast

Detailed domain annotation

Selected features only

Show module domains



Gene details

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