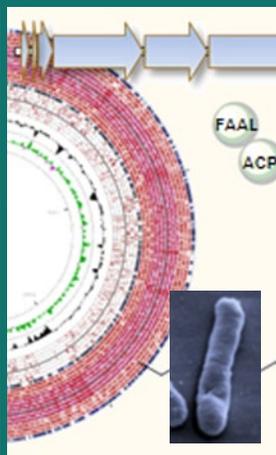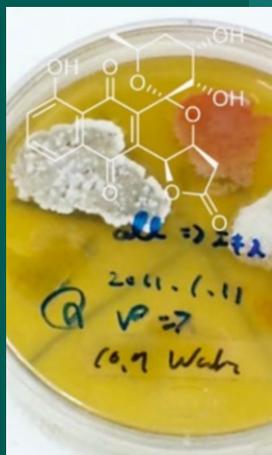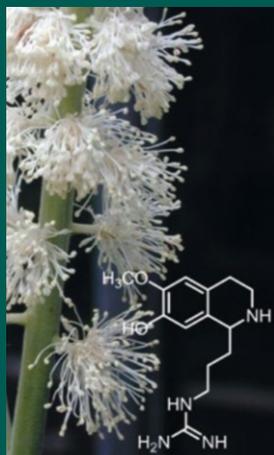## Editors

A. D. Kinghorn · H. Falk · J. Kobayashi

Authors

Feng Qiu, J.B. McAlpine, E.C. Krause, Shao-nong Chen, and G.F. Pauli

N. Gaboriaud-Kolar, S. Nam, and A.-L. Skaltsounis

M. Ishibashi

M. Nett

Progress in the Chemistry
of Organic Natural Products

Founded by L. Zechmeister

Editors:
A.D. Kinghorn, Columbus, OH
H. Falk, Linz
J. Kobayashi, Sapporo

Honorary Editor:
W. Herz, Tallahassee, FL

Editorial Board:
V.M. Dirsch, Vienna
S. Gibbons, London
N.H. Oberlies, Greensboro, NC
Y. Ye, Shanghai

99

Progress in the Chemistry of Organic Natural Products

Authors:
Feng Qiu, J.B. McAlpine, E.C. Krause,
Shao-Nong. Chen, and G.F. Pauli

N. Gaboriaud-Kolar, S. Nam, and A.-L. Skaltsounis

M. Ishibashi

M. Nett

Prof. A. Douglas Kinghorn, College of Pharmacy,
Ohio State University, Columbus, OH, USA

em. Univ.-Prof. Dr. H. Falk, Institut für Organische Chemie,
Johannes-Kepler-Universität, Linz, Austria

Prof. Dr. J. Kobayashi, Graduate School of Pharmaceutical Sciences,
Hokkaido University, Sapporo, Japan

Contents

Listed in PubMed

Contributors

Shao-Nong Chen UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Elizabeth C. Krause UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

James B. McAlpine Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Guido F. Pauli UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Feng Qiu Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Nicolas Gaboriaud-Kolar Department of Pharmacognosy and Natural Products Chemistry, School of Pharmacy, University of Athens, Athens, Greece

Sangkil Nam Molecular Medicine, Beckman Research Institute, City of Hope Comprehensive Cancer Center, Duarte, CA, USA

Alexios-Leandros Skaltsounis Department of Pharmacognosy and Natural Products Chemistry, School of Pharmacy, University of Athens, Athens, Greece

Masami Ishibashi Graduate School of Pharmaceutical Sciences, Chiba University, Chuo-ku, Chiba, Japan

Markus Nett Leibniz Institute for Natural Product Research and Infection Biology, Jena, Germany

About the Authors

Shao-Nong Chen, Ph.D. After completing a B.S. degree in Organic Chemistry from Lanzhou University, China, and an assistantship in the Lanzhou Institute of Chemical Physics (CAS), he pursued his interests in natural products chemistry, obtaining a Ph.D. under the joint mentorship of Drs. Yao-Zu Chen, Lanzhou University, and Han-Dong Sun, Kunming Institute of Botany (KIB, CAS). After a two-year post-Ph.D. training period with Dr. Guo-Wei Qin at Shanghai Institute of Materia Medica (SIMM), he joined Dr. Sydney Hecht's group at the University of Virginia. Since 2000, his research at the UIC/NIH Botanical Center has focused on plant standardization and botanical integrity.
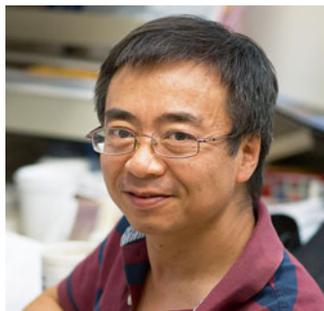
Elizabeth C. Krause, Pharm.D. She received a B.S. with honors in biology from UIC. After a stint as a paralegal with the law firm Vedder Price, she obtained a Pharm.D. degree with the UIC College of Pharmacy. In a variety of positions in the medical and financial sectors, including University Healthsystem Consortium and William Blair, she developed a Drug Utilization Review Program, wrote drug monographs, managed reviews of new products for the healthcare market, and was introduced to Complementary and Alternative Medicine. Recently, she has been melding all of her interests as Program Manager in the UIC/NIH Botanical Center at the University of Illinois at Chicago.

James B. McAlpine, Ph.D. He obtained a Ph.D. from the University of New England, Armidale, NSW, Australia. Postdoctoral work at Northwestern University Medical School, studying the biosynthesis and mode of action of macrolide antibiotics followed. In 1972, he joined Abbott Laboratories and worked on macrolides, aminoglycosides, and quinolones before heading up their natural product discovery group from 1981 to 1996, and he discovered Tiacumicin B, the API of Difficid®. He joined Phytera Inc. as VP of Chemistry in 1996, discovering drugs from manipulated plant cell cultures, and in 2002 joined Ecopia BioSciences as VP of Chemistry and Discovery using genomics to discover novel secondary metabolites. He has authored over 100 research articles and has been named as an inventor on 50 U.S. patents.

Guido F. Pauli, Ph.D. A pharmacist by training, he holds a doctoral degree in pharmacognosy. Currently, a Professor at UIC, Chicago (IL), he is principal investigator and collaborator in interdisciplinary research projects. His main research interests are in bioactivity and metabolomic analysis of complex natural products, herbal dietary supplements, anti-TB drug discovery, and dental applications of natural agents. Scholarly activities involve service on funding agency panels, on pharmacopoeial expert committees, and in professional societies. His portfolio comprises over 130 peer-reviewed journal articles, numerous oral and conference presentations, three book chapters, five patents, as well as journal editorial and board functions.

Feng Qiu, Ph.D. After earning his Bachelor's degree in traditional Chinese medicine in China, he received his Ph.D. degree in Pharmacognosy from the University of Illinois at Chicago (UIC). Currently, a Postdoctoral Fellow at the Samuel Roberts Noble Foundation, Ardmore, Oklahoma, his research focuses on large-scale plant metabolomics. His research interests are in the metabolomic profiling of plants by sophisticated analytical methods, in particular mass spectrometry and NMR spectroscopy, as well as the implementation of computational techniques for integrated data evaluation.
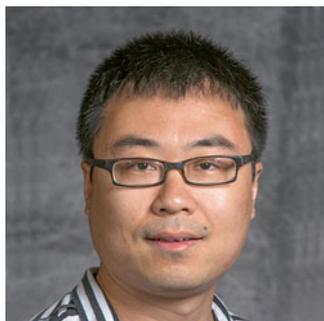
Nicolas Gaboriaud-Kolar is a postdoctoral researcher in the Department of Pharmacognosy and Natural Products Chemistry at the University of Athens. After gaining a Technical University Diploma in Chemistry (2002) and a M.Sc. degree in Chemistry (2004) from the University Paris 12-Val de Marne (France), he completed a Certificate in Further Education in "Chemistry towards Life Sciences" in 2005 at the University Paris Descartes, Paris 5. Interested in natural products chemistry and its applications in thera-peutics, he received his Ph.D. degree in 2008 in Pharmacognosy and Natural Product Chemistry while in the group of Professor François Tillequin. After one year as Temporary Teaching and Research Assistant in the University Paris Descartes, he joined the group of Professor Skaltsounis in 2009. His research focuses on the synthesis and medicinal chemistry of natural products. He has nine peer-reviewed publications and has been named as coinventor of one patent.

Sangkil Nam, Ph.D. is currently working as Director of the High Throughput Screening Core, Beckman Research Institute, City of Hope Comprehensive Cancer Center, which he joined in 2005. His research interest is to identify leads to develop antitumor therapeutic agents. In his efforts on anticancer drug discovery and development at Beckman Research Institute, Dr. Nam has collab-orated with well-known scientists worldwide. He discovered that several novel small molecules, including indirubin analogs, have potential as a new generation of more effective and less toxic therapeutics for human cancer treatment. Prior to joining Beckman Research Institute, Dr. Nam worked as a research scientist at the H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, where he finished postdoctoral training in 2003. Dr. Nam earned his Ph.D. degree in Biochemistry and Molecular Biology, OGI School of Science and Engineering, Oregon Health and Science University in 1998, after fin-ishing the undergraduate program in Chemistry at the University of Kansas in 1994.
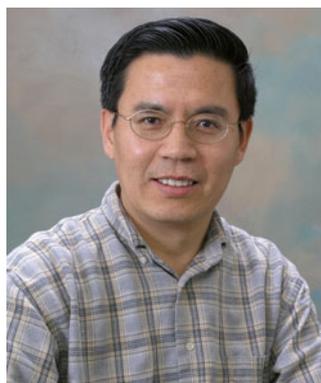
Alexios-Leandros Skaltsounis is Professor and Director of the Department of Pharmacognosy and Natural Product Chemistry, Faculty of Pharmacy, University of Athens, and Dean of the Faculty. He obtained his Ph.D. degree in Pharmacy from Descartes University, Paris V. After six years as an academic staff member in Paris, he joined the University of Athens. At present, he leads a research group of more than 30 researchers, focusing on discovery, development, and exploitation of novel natural products and semisynthetic chemical entities. He is an author of 220 scientific publications and a named coinventor of 12 patents.

Masami Ishibashi is Professor of Natural Products Chemistry at the Graduate School of Pharmaceutical Sciences, Chiba University, Japan. He received his Ph.D. degree in Chemistry at the University of Tokyo in 1985 under the supervision of Takeyoshi Takahashi. After four years of postdoctoral study at the University of Hawaii-Manoa with Richard E. Moore, and the Mitsubishi-Kasei Institute of Life Sciences with Jun'ichi Kobayashi, and the Kitasato Institute with Satoshi Ōmura and Kanki Komiyama, he joined the faculty of Toho University in 1989 and moved to Hokkaido University as an Associate Professor of Pharmacognosy in 1990. He has been a Full Professor at Chiba University since 1997, where he performs research in the field of natural products chemistry. He received the Pharmaceutical Society of Japan Award for Divisional Scientific Promotions in 2007 and the Sumiki-Umezawa Memorial Award from the Japan Antibiotic Research Association in 2012. He has been a Section Editor-in-Chief of *Chemical and Pharmaceutical Bulletin* since 2013.

Markus Nett was born in 1977, studied pharmacy at the University of Bonn. Following his licensure, he graduated from the research group of Prof. Dr. Gabriele König (University of Bonn) and obtained his Ph.D. in pharmaceutical sciences with a project on the isolation and structure elucidation of bioactive natural products from gliding bacteria. In 2007, he performed postdoctoral research in the group of Professor Bradley S. Moore (University of California at San Diego) and focused on the engineering of

biosynthesis pathways in marine microorganisms. In 2008, he became a DAAD fellow at the Scripps Institution of Oceanography in San Diego. Since 2009, he has been head of a junior research group at the Leibniz Institute for Natural Product Research and Infection Biology (Hans-Knöll-Institute) in Jena, where he investigates the potential of predatory bacteria and plant pathogens to produce new antibiotics. His research interests are in exploring and exploiting microbial genomes to identify new biosynthetic pathways and natural products for drug discovery and development.

Pharmacognosy of Black Cohosh: The Phytochemical and Biological Profile of a Major Botanical Dietary Supplement

Feng Qiu, James B. McAlpine, Elizabeth C. Krause, Shao-Nong Chen, and *Guido F. Pauli*

Contents

F. Qiu • J.B. McAlpine
Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy,
University of Illinois at Chicago, 833 South Wood Street, Chicago, IL 60612, USA
e-mail: fqiu2@uic.edu; mcalpine@uic.edu

E.C. Krause • S.-N. Chen • G.F. Pauli (✉)
UIC/NIH Center for Botanical Dietary Supplements Research, Department of Medicinal
Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago,
833 South Wood Street, Chicago, IL 60612, USA
e-mail: krause@uic.edu; sc4sa@uic.edu; gfp@uic.edu

1 Introduction

The term "Black Cohosh" has been used in both the scientific and commercial literature with a wide variety of meanings including the source plant, portions thereof, and plant extracts. In this chapter, the authors will attempt to use this term and related terms including the Latin binomial rigorously, with the following meanings:

Actaea racemosa (*A. racemosa*) strictly for the source **plant**;
Black Cohosh (BC) for the **underground parts** (roots and rhizomes) of the plant;
Black Cohosh Extract (BCE) for an **extract of BC**.

Difficulties in defining the exact nature of each of these, particularly the last term, will be discussed later in sections covering botany, verification, adulteration, quality control, naming, and pharmacology.

   Actaea racemosa L. (syn. *Cimicifuga racemosa*) [L.] Nutt.) is a species of flowering plant of the family Ranunculaceae (Plate 1). It is an endemic native to North America and distributed mainly from the south of Ontario to central Georgia, and west to Missouri and Arkansas. The roots and rhizomes (Plate 2) have been used medicinally by Native Americans since pre-Columbian times, to treat malaise, gynecological disorders, kidney disorders, malaria, rheumatism, and sore throats (*1*). Following the appearance of BC as "Black Snakeroot" in the U.S. Pharmacopoeia in 1820 (*2*), BCEs have gained popularity in contemporary use as herbal dietary supplements to alleviate women's menopausal symptoms. A recent report released by the American Botanical Council states that BCE is one of the ten top-selling herbal dietary supplements in the food, drug, and mass market channel in the United States in 2011, with sales of over $10 million (*3*, *4*). The widespread use of BC as a dietary supplement has also led to an increasing call for scientific evidence of its quality, safety, and efficacy.

   In 1999, the University of Illinois at Chicago (UIC)/National Institutes of Health (NIH) Center for Botanical Dietary Supplements Research initiated a systematic study of the entire subject (*A. racemosa*, BC, and BCEs) and since then has dedicated more than 10 years of research to enhance knowledge of their botany, chemistry, and biology. The present contribution summarizes and analyzes the progress of these studies and those by others, including the phytochemical investigation, fingerprint profiling, biological and pharmacological evaluation, and clinical trials.

   Emphasis is placed on the increased recognition of the plant's chemical diversity, which forms the basis of the understanding of its biological and clinical potential. The phytochemistry of *A. racemosa* is first reviewed with a focus on the potential bioactive components including cycloartane triterpenes, cimicifugic acids, phenolics, and nitrogen-containing compounds. For each of these major types of constituents, a compilation of the structures and nomenclature of the individual compounds, known to date, is provided. Isolation, purification, and structural characterization methods for these compounds are also discussed. The subsequent section provides

Plate 1 Aerial parts of the title plant, *Actaea racemosa* L.: entire plant (*bottom*), inflorescence (*middle*), flower (*top*), in front of its leaves

an overview of the chemical fingerprinting of *A. racemosa*, which has evolved as an important approach used for the metabolomic profiling, botanical identification, and quality standardization. The section on pharmacology of BCEs provides a concise yet relatively comprehensive literature survey of the major bioactivities of both crude extracts and pure compounds, such as "estrogenic" activity, anticancer activity, stress relief, and prevention of bone loss. Furthermore, a detailed description is

Plate 2 Dried roots and rhizomes of *Actaea racemosa* L.: entire rootstock (*top*) and detail (*bottom*)

given of the reports of BCE-associated hepatotoxicity with the inherent safety concern. The final section briefly summarizes and discusses the key points of the clinical research of BCEs conducted in the last 20 years, which provides the reader with an overview of the progress and challenges of the safety and efficacy evaluation of BCEs and related botanical preparations in clinical settings.

By assuming an up-to-date perspective of the various aspects of this widely used botanical, it can be demonstrated that the chemical diversity of *A. racemosa* is far broader than what has typically been considered in experimental (*in vitro*, *in vivo*, and clinical) approaches taken to study its biological significance. By placing

emphasis on the underlying chemistry and viewing BCE preparations as complex mixtures of pharmacologically active agents with highly diverse biological targets, the present work is intended to inspire future research aimed at recently characterized phytoconstituents and those still to be discovered, and the evaluation of their biological potential.

2 Phytochemistry

2.1 Cycloartane Triterpenes

2.1.1 Structural Diversity

Cycloartane triterpenes are the most well studied components of *Actaea racemosa*. These compounds possess complex chemical structures as well as many potential bioactivities. Since the 1950s, more than 40 cycloartane triterpenes have been isolated and identified from *A. racemosa* (*5*). A few were identified in the aerial parts (*6*). These triterpenes predominantly occur as 3-*O*-β-D-xylopyranosides or, less commonly, as 3-*O*-α-L-arabinopyranosides. Based on the structures of their aglycone side chains, they are classified into several subgroups, as shown in Fig. 1. The structural diversity of *Actaea* triterpenes is also contributed to by substitution of the core with functional groups such as a double bond at C-7, C-8, hydroxy groups at C-1, C-12 and/or C-15, and acetoxy groups at C-12 and/or C-15.

Tables 1–9 list, by type, all cycloartane triterpenes identified in *A. racemosa* prior to 2013. A bar graph of the numbers of the different types of cycloartane triterpenes identified in *A. racemosa* (see Fig. 1) shows that the cimigenol and acteol types are the major triterpene constituents, in terms of both the number and abundance of these compounds. The hydroshengmanol- and 23-*O*-acetylshengmanol-type triterpenes are commonly found in *A. simplex*, and a few of them have also been found in *A. racemosa*. The other types of *Actaea* triterpenes, such as the dahurinols, cimicidanols, and cimicidols, have been found only rarely in *A. racemosa*, possibly due to their non-occurrence or significantly lower abundance. This represents a characteristic biosynthetic profile of triterpenes for *A. racemosa*.

Despite the very interesting and complex structures of many of these triterpenes, especially those containing 3-, 5-, and 6-membered oxygen heterocycles, to the best of our knowledge no synthesis or semisynthesis approaches have been undertaken. The synthesis of these molecules would appear to present a formidable challenge to the synthesis chemistry community.
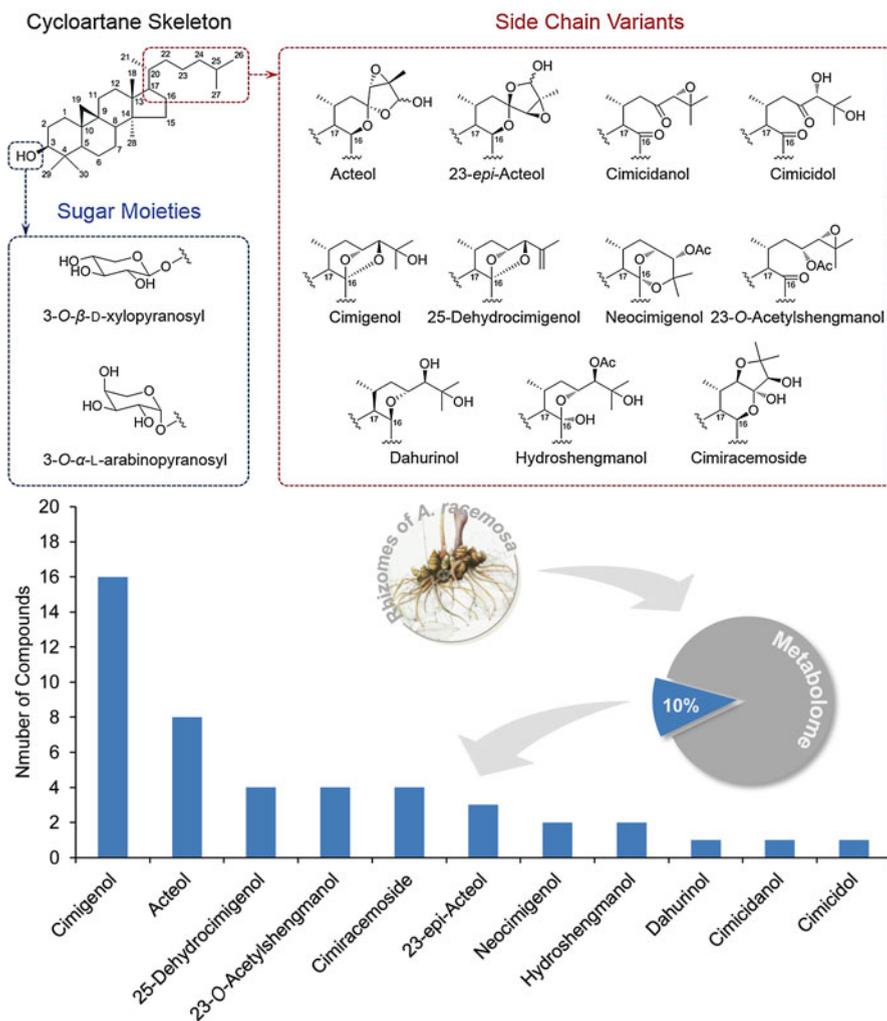
Fig. 1 The structural types (*top*) and numbers (*bottom*) of cycloartane triterpenes identified in *A. racemosa*. As all structures reported so far exhibit the same configuration *alpha* at carbon C-17, it is not drawn in the subsequent structural formulas in the present work

Table 1 The cimigenol-type triterpenes identified in *A. racemosa*

Cimigenol (Acta-16,23;16,24-binoxol)

No.	Common name	New systematic name	R^1	R^2	R^3	R^4	Ref.
1	Cimigenol-3-*O*-β-D-xylopyranoside (Cimigenoside)	(15*R*)-15,25-Dihydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	xyl	H	Me	OH	(10)
2	Cimigenol-3-*O*-α-L-arabinopyranoside (Cimiracemoside C)	3-*O*-α-L-Arabinopyranosyl-(15*R*)-15,25-dihydroxy-acta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	ara	H	Me	OH	(77)
3	25-*O*-Methylcimigenol-3-*O*-β-D-xylopyranoside	(15*R*)-15-Hydroxy-25-methoxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	xyl	H	Me	OMe	(138)
4	25-*O*-Methylcimigenol-3-*O*-α-L-arabinopyranoside	3-*O*-α-L-Arabinopyranosyl-(15*R*)-15-hydroxy-25-methoxyacta-(16*S*,23*R*,24*S*)-binoxoside	ara	H	Me	OMe	(76)
5	25-*O*-Acetylcimigenol-3-*O*-β-D-xylopyranoside	25-Acetoxy-(15*R*)-15-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	xyl	H	Me	OAc	(10)
6	25-*O*-Acetylcimigenol-3-*O*-α-L-arabinopyranoside	25-Acetoxy-3-*O*-α-L-arabinopyranosyl-(15*R*)-15-hydroxyacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	ara	H	Me	OAc	(10)
7	25-Chlorodeoxycimigenol-3-*O*-β-D -xylopyranoside	25-Chloro-(15*R*)-15-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside	xyl	H	Me	Cl	(14)

(continued)

Table 1 (continued)

#	Name					Ref.
8	12β-Hydroxycimigenol-3-O-β-D-xylopyranoside	(12R,15R)-12,15-Dihydroxy-3-O-β-D-xylopyranosylacta-(16S,23R,24S)-16,23;16,24-binoxoside	xyl	OH	Me	OH	(138)
9	12β-Hydroxycimigenol-3-O-α-L-arabinopyranoside	3-O-α-L-Arabinopyranosyl-(12R,15R)-12,15-dihydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	ara	OH	Me	OH	(77)
10	12β-Acetoxycimigenol-3-O-β-D-xylopyranoside	(12R)-12-Acetoxy-(15R)-15-hydroxy-3-O-β-D-xylopyranosylacta-(16S,23R,24S)-16,23;16,24-binoxoside	xyl	OAc	Me	OH	(77)
11	12β-Acetoxycimigenol-3-O-α-L-arabinopyranoside (Cimiracemoside D)	(12R)-12-Acetoxy-3-O-α-L-arabinopyranosyl-(15R)-15-hydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	ara	OAc	Me	OH	(77, 138)
12	25-O-Acetyl-12β-hydroxy-cimigenol-3-O-α-L-arabinopyranoside	25-Acetoxy-3-O-α-L-arabinopyranosyl-(12R)-12-hydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	ara	OH	Me	OAc	(138)
13	12β,21-Dihydroxycimigenol-3-O-α-L-arabinopyranoside	3-O-α-L-Arabinopyranosyl-(12R)-12,21,25-trihydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	ara	OH	CH$_2$OH	OH	(138)
14	21-Hydroxycimigenol-3-O-β-D-xylopyranoside	21,25-Dihydroxy-3-O-β-D-xylopyranosylacta-(16S,23R,24S)-16,23;16,24-binoxoside	xyl	H	CH$_2$OH	OH	(77)
15	21-Hydroxycimigenol-3-O-α-L-arabinopyranoside	3-O-α-L-Arabinopyranosyl-21,25-dihydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	ara	H	CH$_2$OH	OH	(77)
16	Cimigenol	25-Hydroxyacta-(16S,23R,24S)-16,23;16,24-binoxoside	H	H	H	OH	(86)

Table 2 The 25-dehydrocimigenol-type triterpenes identified in *A. racemosa*

25-Dehydrocimigenol (25-Dehydroacta-16,23;16,24-binoxol)

No.	Common name	New systematic name	R¹	R²	Ref.
17	25-Dehydrocimigenol-3-*O*-β-D-xylopyranoside	25-Dehydro-(15*R*)-15-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside	xyl	H	(*10*)
18	25-Dehydrocimigenol-3-*O*-α-L-arabinopyranoside	3-*O*-α-L-Arabinopyranosyl-25-dehydro-(15*R*)-15-hydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside	ara	H	(*10*)
19	12β-Acetoxy-25-dehydrocimigenol-3-*O*-β-D-xylopyranoside (Cimiracemoside K)	(12*R*)-12-Acetoxy-25-dehydro-(15*R*)-15-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside	xyl	OAc	(*10*)
20	12β-Acetoxy-25-dehydrocimigenol-3-*O*-α-L-arabinopyranoside (Cimiracemoside J)	(12*R*)-12-Acetoxy-3-*O*-α-L-arabinopyranosyl-25-dehydro-(15*R*)-15-hydroxyacta-(16*S*,23*R*,24*R*)-16,23;16,24-binoxoside	ara	OAc	(*10*)

Table 3 The acteol-type triterpenes identified in *A. racemosa*

Acteol (Acta-16,23;23,26-binoxol)

No.	Common name	New systematic name	R^1	R^2	R^3	7,8	Ref.
21	12β-Acetoxy-26-deoxyacteol	(12R)-12-Acetoxy-(24R,25S)-24,25-epoxyacta-(16S,23S)-16,23;23,26-binoxol	H	OAc	H	2H	(139)
22	12β-Acetoxy-26-deoxyacteol-3-O-β-D-xylopyranoside (26-Deoxyactein)	(12R)-12-Acetoxy-(24R,25S)-24,25-epoxy-3-O-β-D-xylopyranosylacta-(16S,23S)-16,23;23,26-binoxoside	xyl	OAc	H	2H	(140)
23	12β-Acetoxyacteol	(12R)-12-Acetoxy-(24R,25S)-24,25-epoxy-(26R/S)-26-hydroxyacta-(16S,23R)-16,23;23,26-binoxol	H	OAc	OH	2H	(141)
24	12β-Acetoxyacteol-3-O-β-D-xylopyranoside (Actein)	(12R)-12-Acetoxy-(24R,25S)-24,25-epoxy-(26R/S)-26-hydroxy-3-O-β-D-xylopyranosylacta-(16S,23R)-16,23;23,26-binoxoside	xyl	OAc	OH	2H	(140)
25	12β-Acetoxyacteol-3-O-β-D-(2'-O-acetyl)-xylopyranoside	(12R)-12-Acetoxy-3-O-β-D-(2'-O-acetyl)xylopyranosyl-(24R,25S)-24,25-epoxy-(26R/S)-26-hydroxyacta-(16S,23R)-16,23;23,26-binoxoside	2'-OAc-xyl	OAc	OH	2H	(10)
26	12β-Acetoxyacteol-3-O-β-D-(4'-O-acetyl)-xylopyranoside (Cimiracemoside O)	(12R)-12-Acetoxy-3-O-β-D-(4'-O-acetyl)-xylopyranosyl-(24R,25S)-24,25-epoxy-(26R/S)-26-hydroxyacta-(16S,23R)-16,23;23,26-binoxoside	4'-OAc-xyl	OAc	OH	2H	(10)
27	12β-Acetoxy-26-ketoacteol-3-O-β-D-xylopyranoside (Cimiracemoside P)	(12R)-12-Acetoxy-(24R,25S)-24,25-epoxy-26-keto-3-O-β-D-xylopyranosylacta-(16S,23R)-16,23;23,26-binoxoside	xyl	OAc	=O	2H	(10)
28	12β-Acetoxy-7,8-dedihydro-26-deoxyacteol-3-O-β-D-xylopyranoside (26-Deoxycimicifugoside)	(12R)-12-Acetoxy-7,8-dedihydro-(24R,25R)-24,25-epoxy-3-O-β-D-xylopyranosylacta-(16S,23S)-16,23;23,26-binoxoside	xyl	OAc	H	$\Delta^{7,8}$	(10)
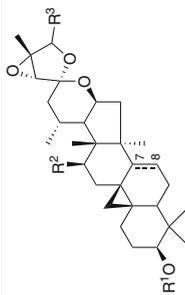
Table 4 The 23-*epi*-acetol-type triterpenes identified in *A. racemosa*

Acteol (Acta-16,23;23,26-binoxol)

No.	Common name	New systematic name	R¹	R²	7,8	Ref.
29	12β-Acetoxy-23-*epi*-26-deoxyacteol-3-*O*-β-D-xylopyranoside (23-*epi*-26-Deoxyactein)	(12*R*)-12-Acetoxy-(24*R*,25*R*)-24,25-epoxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside	xyl	OAc	2H	(*10*)
30	12β-Acetoxy-23-*epi*-26-deoxyacteol-3-*O*-α-L-arabinopyranoside (Cimiracemoside N)	(12*R*)-12-Acetoxy-3-*O*-α-L-arabinopyranosyl-(24*R*,25*R*)-24,25-epoxyacta-(16*S*,23*R*)-16,23;23,26-binoxoside	ara	OAc	2H	(*10*)
31	7,8-Dedihydroacteol-3-*O*-β-D-xylopyranoside (Cimiracemoside I)	(24*R*,25*R*)-24,25-epoxy-7,8-dedihydro-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside	xyl	H	Δ^{7,8}	(*10*)

Table 5 The 23-*O*-acetylshengmanol-type triterpenes identified in *A. racemosa*
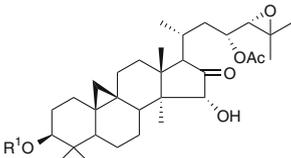
23-*O*-Acetylshengmanol (23-Acetoxy-24,25-epoxyactanol)

No.	Common name	New systematic name	R¹	Ref.
32	23-*O*-Acetylshengmanol-3-*O*-β-D-xylopyranoside	(23*R*)-23-Acetoxy-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxo-3-*O*-β-D-xylopyranosylactanoside	xyl	(*76*)
33	23-*O*-Acetylshengmanol-3-*O*-α-L-arabinopyranoside	(23*R*)-23-Acetoxy-3-*O*-α-L-arabinopyranosyl-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside	ara	(*10*)
34	23-*O*-Acetylshengmanol-3-*O*-β-D-(4′-*O*-acetyl)-xylopyranoside (Cimiracemoside M)	(23*R*)-23-Acetoxy-3-*O*-β-D-(4′-*O*-acetyl)xylopyranosyl-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside	4′-OAc-xyl	(*10*)
35	23-*O*-Acetylshengmanol-3-*O*-α-L-(4′-*O*-acetyl)-arabinopyranoside (Cimiracemoside L)	(23*R*)-23-Acetoxy-3-*O*-α-L-(4′-*O*-acetyl)arabinopyranosyl-(24*S*)-24,25-epoxy-(15*R*)-15-hydroxy-16-oxoactanoside	4′-OAc-ara	(*10*)

Table 6 The cimiracemoside-type triterpenes identified in *A. racemosa*
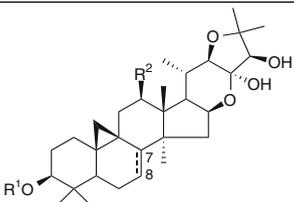
Cimiracemoside (Acta-16,23;22,25-binoxoside)

No.	Common name	New systematic name	R¹	R²	7,8	Ref.
36	N/A	(12*R*)-12-Acetoxy-3-*O*-α-ʟ-arabinopyranosyl-(23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxoside	ara	OAc	2H	(*138*)
37	Cimiaceroside A	7,8-Didehydro-(23*R*,24*R*)-23,24-dihydroxy-3-*O*-β-ᴅ-xylopyranosylacta-(16*S*,22*R*)-16,23;22,25-binoxoside	xyl	H	Δ⁷,⁸	(*144*)
38	Cimiracemoside F	(12*R*)-12-Acetoxy-7,8-Didehydro-(23*R*,24*R*)-23,24-dihydroxy-3-*O*-β-ᴅ-xylopyranosylacta-(16*S*,22*R*)-16,23;22,25-binoxoside	xyl	OAc	Δ⁷,⁸	(*77*) (*76*)
39	Cimiracemoside G	(12*R*)-12-Acetoxy-3-*O*-α-ʟ-arabinopyranosyl-7,8-didehydro-(23*R*,24*R*)-23,24-dihydroxyacta-(16*S*,22*R*)-16,23;22,25-binoxoside	ara	OAc	Δ⁷,⁸	(*77*)
40	Cimiaceroside B	(23*R*,24*R*)-23,24-dihydroxy-3-*O*-β-ᴅ-xylopyranosylacta-(16*S*,22*R*)-16,23;22,25-binoxoside	xyl	H	2H	(*144*)

Table 7 The hydroshengmanol-type triterpenes identified in *A. racemosa*
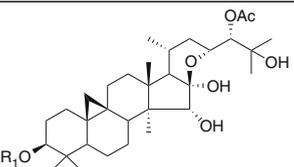
Hydroshengmanol (Acta-16,23-monoxol)

No.	Common name	New systematic name	R¹	Ref.
41	24-*O*-Acetylhydroshengmanol-3-*O*-β-ᴅ-xylopyranoside	(24*S*)-24-Acetoxy-(15*R*,16*R*)-15,16,25-trihydroxy-3-*O*-β-ᴅ-xylopyranosylacta-(23*S*)-16,23-monoxoside	xyl	(*10*)
42	24-*O*-Acetylhydroshengmanol-3-*O*-α-ʟ-arabinopyranoside	(24*S*)-24-Acetoxy-3-*O*-α-ʟ-arabinopyranosyl-(15*R*,16*R*)-15,16,25-trihydroxyacta-(23*S*)-16,23-monoxoside	ara	(*142*)

Table 8 The neocimigenol-type triterpenes identified in *A. racemosa*.
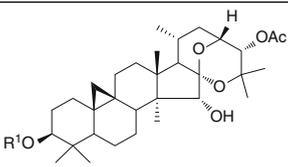
Neocimigenol (Acta-16,23;16,25-binoxol)

No.	Common name	New systematic name	R¹	Ref.
43	Neocimicigenoside A	(24*S*)-24-Acetoxy-(15*R*)-15-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*)-16,23;16,25-binoxoside	xyl	(*7*)
44	Neocimicigenoside B	(24*S*)-24-Acetoxy-3-*O*-α-L-arabinopyranosly-(15*R*)-15-hydroxyacta-(16*S*,23*R*)-16,23;16,25-binoxoside	ara	(*7*)

Table 9 The isodahurinol-, cimicidanol-, and cimicidol-type triterpenes identified in *A. racemosa*

No.	Structure	Common name	New systematic name	Ref.
45	Isodahurinol	Cimiracemoside E	(24*S*)-24-Acetoxy-25-hydroxy-15-oxo-3-*O*-β-D-xylopyranosyl-acta-(16*R*,23*R*)-16,13-monoxoside	(*77*)
46	Cimicidanol	Cimicidanol-3-*O*-β-D-xylopyranoside (Cimicifugoside H-1)	7,8-Didehydro-(11*S*,15*R*)-11,15-dihydroxy-(24*R*)-24,25-epoxy16,23-dioxo-3-*O*-β-D-xylopyranosylactanoside	(*10*)
47	Cimicidol	Cimcicidol-3-*O*-β-D-xylopyranoside (Cimicifugoside H-2)	7,8-Didehydro-(11*S*)-11-hydroxy-(24*R*)-24,25-epoxy-16,23-dioxo-3-*O*-β-D-xylopyranosylactanoside	(*10*)
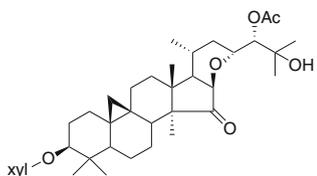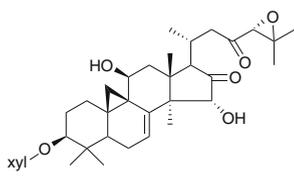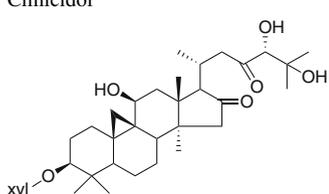
2.1.2 Naming System

Previously, *Actaea/Cimicifuga* triterpenes were given a wide variety of trivial names based on the species, traditional Chinese medicinal use, and only occasionally following some structural features. A comprehensive list of the names of aglycones includes acetol, cimiaceroside, cimicidanol, cimicidol, cimifugoside, cimigenol, cimiracemoside, dahurinol, hydroshengmanol, neocimigenols, neocimicigenosides, and 23-*O*-acetylshengmanol (*5*, *7*). These trivial names, at best, provide clues as to the origin of the compound, but the similarity of names gives no indication of any similarity of the structures. An example of the enormous confusion created by the trivial names given to the *Actaea/Cimicifuga* triterpenes may be demonstrated by considering cimiracemosides F (**38**), M (**34**), and P (**27**): these compounds have completely different ring systems and also differ from one another in the sites of oxygenation at C-12, C-15, C-16, C-22, C-23, and C-26 (see Fig. 2). The reverse situation occurs in the pair,

Cimiracemoside F

Cimiracemoside M

Cimiracemoside P
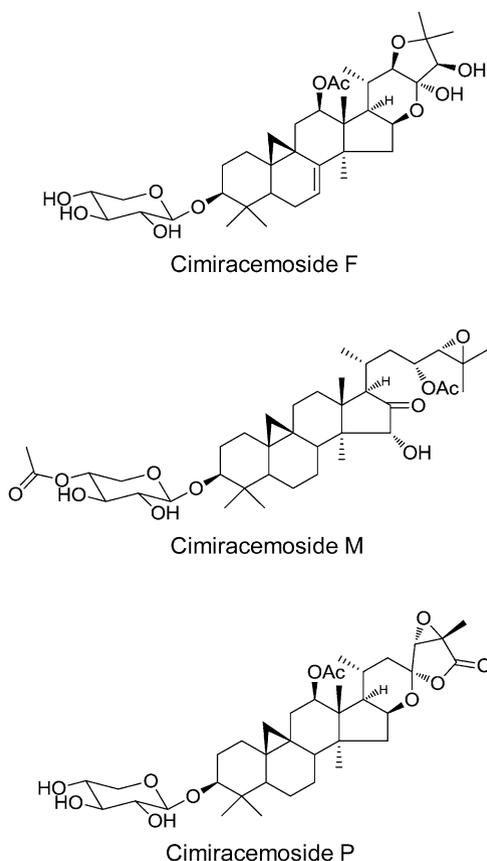
Fig. 2 Three *Actaea* triterpenes that have the same trivial root name but completely different ring systems in their side chains

neocimigenol and neocimicigenoside A (**43**). While the former has the identical agly-cone moiety to the latter, its name varies in a very subtle and rather confusing way ("cimi" *vs.* "cimici"). Considering that the triterpene structures alone differ in very subtle ways, it becomes essentially impossible to differentiate between subtle structural and subtle naming changes, especially if the naming schemes lack a clear rationale.

To address this problem, recently a new systematic naming system was estab-lished, which enables the deduction of all known *Actaea* triterpene structures, as well as those of congeners yet to be discovered. All that is required is a knowledge of the basic cycloartane skeleton (*8*). As shown in Fig. 1, all of the known cycloar-tane triterpenes from *A. racemosa* fall into only a few basic structural skeletons. As far as C-20 to C-27 are concerned, there are acyclic compounds in which these carbons have no connections between themselves other than the basic carbon chain. Then, there are other compounds in which some of these carbons are involved in one or two rings, usually formed by ether or acetal oxygens, often connecting to C-16. On the basis of these structural features, the new system names the acyclic aglycones as *actanols*, those with a single oxygen bridge forming a further ring as *actamonoxols*, and those with two oxygen-containing rings *actabinoxols*. (Note: oxiranes are not counted in determining the mono- or bi-oxonol name; see Fig. 3).
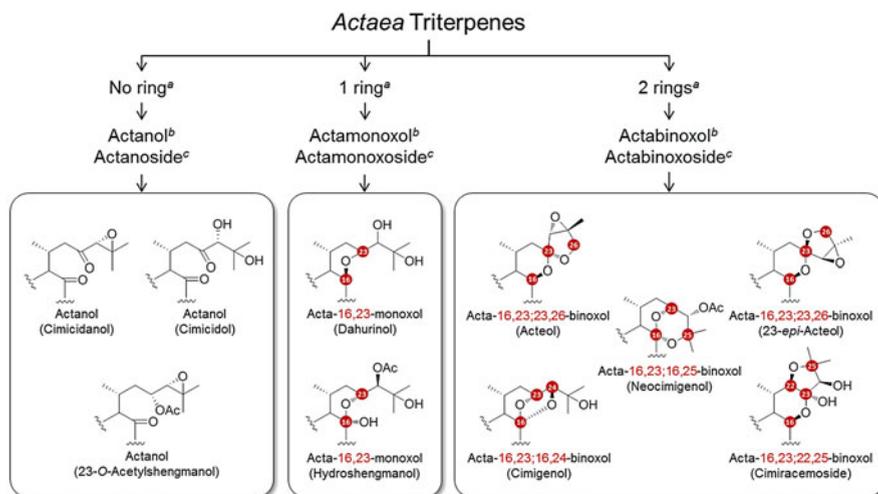
Fig. 3 Naming system for *Actaea* triterpenes established by *Qiu et al.* (*8*). [a]Number of rings in the aglycone side chain. [b]Aglycone. [c]Glycoside

These names all include the 3β-hydroxy group. Where this group is modified to a glycosidic linkage, the suffix would be "-oside", *e.g.* actabinoxoside. All of the substituents and other structural modifications need to be fixed using standard chemical nomenclature, with prefixes arranged in alphabetical order. The configura-tion of these triterpenes is designated *via* the *Cahn-Ingold-Prelog* (*CIP*) system rather than the simpler α/β system used commonly in steroid and terpenoid nomen-clature, because the α/β nomenclature fails in bicyclic caged rings that occur in many of the actabinoxols.

2.1.3   Isolation Techniques

Actaea triterpenes are difficult to separate and purify because they occur as mixtures of natural product congeners with highly similar chemical properties. Repeated isolation techniques reported in the literature include gravity and low-pressure liquid chromatography (LPLC), medium-pressure liquid chromatography (MPLC), vacuum-liquid chromatography (VLC), high-speed countercurrent chromatography (HSCCC), and high-performance liquid chromatography (HPLC). Commonly, three to eight steps of fractionation are needed to obtain the compounds with sufficient purity for structural elucidation. It should be noted that from the experience of the authors, the purity level required for the elucidation of structure is relatively low. Moreover, reference materials of *Actaea/Cimicifuga* triterpenes commonly exhibit significant residual complexity (*8*) and require thorough analysis, especially prior to being used for any meaningful biological evaluation.

The isolation workflow of *Actaea* triterpenes usually starts with liquid partitions of the methanol extract of plant materials using a series of organic solvents. This results in triterpene-enriched partitions in phases of medium polarity, such as $CHCl_3$ and *n*-BuOH. Subsequent procedures include repetitive fractionation by normal-phase (*e.g.* silica gel), reversed-phase (*e.g.* C_8 and C_{18} silica gel), and, occasionally, Sephadex (*e.g.* LH-20) column chromatography. Analytical thin-layer chromatography (TLC) has been used to optimize the solvent conditions for the preparative separation on the column chromatography, and using the G.U.E.S.S method (*9*) to choose an HSCCC solvent system. Isolation of pure compounds can take up to eight fractionation steps even when the conditions are carefully optimized (*10*). For the resolution of critical pairs of triterpenes, preparative RP-HPLC not only provides optimized resolution, but also reduces sample loss and overall workload. Therefore, HPLC has been used frequently as in the last purification step for these compounds. As the *Actaea* triterpenes exhibit very weak UV absorption, online ELSD or offline TLC analysis are needed to optimize and monitor the HPLC separation process. Recently, *Sezai, Cizek et al.* have reported a fast and convenient method for the isolation of triterpenes from *A. racemosa* by HSCCC/ELSD (*11*). The optimized solvent system, consisting of *n*-hexane–acetone–ethyl acetate–*iso*-propanol–ethanol–water (7:2:4:2:1:4, *v/v/v/v/v*), enabled a single step of separation of four *Actaea* triterpenes with high yield and purity from an enriched fraction. This demonstrates the high resolution power of CCC, provided the target analytes elute in the "sweet spot" of the chromatogram (*12*).

2.1.4   Structural Elucidation

Most modern approaches for the structural elucidation of the *Actaea* triterpenes involve the use of high-resolution NMR spectroscopy and mass spectrometry techniques, whereby a combination of both has proven to be the most efficient way for the characterization of these molecules. At the same time, X-ray diffraction is vital

for the establishment of absolute stereochemical assignments, which, due to the lack of chromophores among these compounds, are otherwise unachievable with the usual chiroptical methods.

2.1.4.1 Mass Spectrometry

High-resolution mass spectrometry (HR-MS) has been commonly used for the determination of the molecular formula of *Actaea* triterpenes. Soft-ionization techniques, such as fast-atom bombardment (FAB) and electrospray ionization (EI), have made a major contribution. From the early to mid-1990s, positive-mode HR-FAB-MS was frequently employed. This method uses a liquid matrix, which facilitates production of molecular ions of the solutes. In positive HR-FAB-MS, the quasi-molecular cations, such as $[M+H]^+$, $[M+Na]^+$, and $[M-OH]^+$, are often produced. Furthermore, the peaks in FAB-mass spectra have been used occasionally for the determination of the presence of sugar moieties of *Actaea* triterpenes. For example, the peak $[M+H-132]^+$ indicates the elimination of one pentosyl group (*13*). Since the late 1990s, positive HR-ESI-MS has been also applied for the structural characterization of *Actaea* triterpenes. Similarly, the molecular formulas are readily determined based on the molecular ion peaks. MS has a particular advantage in the identification of halogenated compounds, where the characteristic isotope pattern can be observed in the mass spectra, whereas the NMR spectra provide no direct evidence for the presence of chlorine. Using HR-ESI-MS, *Chen et al.* identified a rare chlorine-containing triterpene, 25-chloro-25-deoxycimigenol-3-*O*-*β*-D-xylopyranoside (**7**), from *A. racemosa* as an artifact of the isolation procedure, and, thus, established a potential pathway for *in vitro/in vivo* chemical diversification of the parent triterpenes with oxirane partial structures (*14*).

2.1.4.2 Nuclear Magnetic Resonance

One-dimensional/two-dimensional (1D/2D) NMR spectroscopic procedures are the most important and useful techniques for the structural elucidation of *Actaea* triterpenes. Due to its solubility properties and achieved signal dispersion, pyridine-d_5 is used widely as the "standard" NMR solvent for *Actaea* triterpenes. Initial ^{1}H NMR analysis typically reveals five to eight singlet methyl signals, one methyl doublet signal (H-21), and, in the range 0.3–0.9 ppm, a pair of doublet signals of cyclopropane protons (H-19). The presence and type of sugar moieties are also easily identified by the signals at lower field (3.8–5.0 ppm). Owing to the significant peak overlap in the upfield portion of the ^{1}H NMR spectra, it is difficult to assign some of the methine and methylene protons, *e.g.* H_2-1, H_2-2, H-5, H_2-6, H_2-7, H-8, H_2-11, H_2-12, and H_2-15. Thus, 2D-NMR spectra are required to aid in the assignment of these proton resonances. The HMBC spectra are particularly useful due to the $^{2–3}J_{C,H}$ correlations between several methyl protons and nearby methine or methylene carbons.
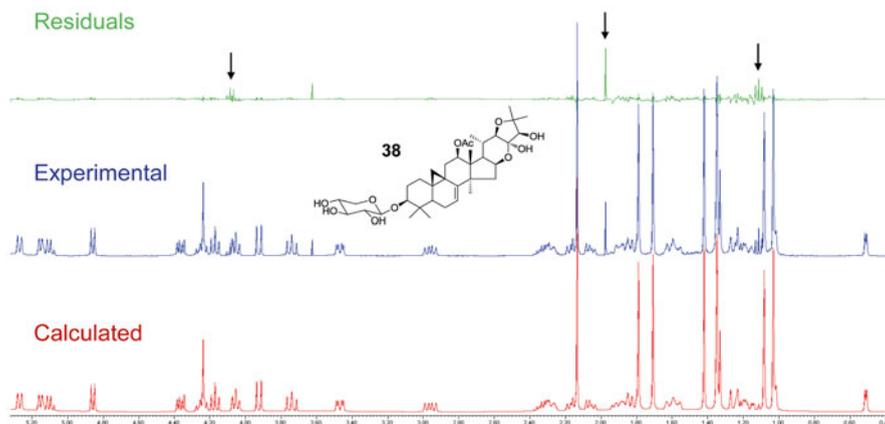
Fig. 4 Quantum mechanical interpretation of the ^{1}H NMR spectrum of cimiracemoside F (**38**) in pyridine-d_5 at 400 MHz by ^{1}H iterative Full Spin Analysis (HiFSA) using PERCH software. The difference spectrum (residual) indicates the excellent agreement between the experimental and calculated spectra. The only major residual signals denoted by ↓ belong to EtOAc, which was identified as the residual solvent in this sample

Upon the determination of these δ_C values, the δ_H values of methylenes and methines can be assigned from the HSQC spectra. In addition, the HSQC and COSY spectra may be useful to deduce the $J_{H,H}$ values. However, significant peak overlap and higher order spin systems often still severely limit the identification of distinct multiplicity patterns. This problem can be solved by using spectroscopic simulation *via* ^{1}H iterative Full Spin Analysis (HiFSA) (*15*, *16*). Figure 4 shows an example of such a full spin analysis of an *Actaea* triterpene, which involves the use of PERCH software. Upon conformational analysis and prediction of basic NMR parameters (δ_H, $J_{H,H}$) using a 3D model of the analyte, the ^{1}H NMR spectrum was simulated using the predicted parameters. Subsequently, iterative and systematic variation of the $\delta_H/J_{H,H}$ values, simulation of the ^{1}H NMR spectra, and comparison with the experimental spectrum were performed. Finally, the calculated NMR parameters were optimized using Total-Line-Shape (TLS) iteration until the difference between the simulated and experimental spectra (residual) was minimal (typically < 0.1% RMSD).

Previously, the relative configuration of *Actaea* triterpenes has usually been studied by NOESY experiments. With the completion of the full spin analysis (HiFSA), the relative configuration of the protons can be determined readily from the calculated $J_{H,H}$ values. For example, the *axial* and *equatorial* positions of the protons can be determined based on the fact that *axial-axial* couplings are large (~13 Hz), while *axial-equatorial* and *equatorial-equatorial* couplings are much

smaller (~3–5 Hz). The availability of a complete set of *J* values from the HiFSA provides a comprehensive definition of the complex *J*-coupling network in the triterpenes and, thereby, leads to an unequivocal assignment of the relative stereo-chemistry of most of the triterpene skeleton from only a 1D ^{1}H NMR spectrum. Notable exceptions are parts of the O-heterocycles (E/F rings), which can escape this determination by virtue of lack of *J*-coupling. These situations have to be addressed with N*O*E measurements or X-ray diffraction.

2.1.4.3 X-Ray Diffraction

Single crystal X-ray diffraction is by far the most powerful experimental method for the characterization of atomic arrangements in molecules. It provides accurate data concerning both the configuration and conformation of compounds, such as precise atomic coordinates, geometries, and crystal packing in the solid phase. However, X-ray diffraction has not been applied as a routine technique for the conformation studies of *Actaea* triterpenes, because it is challenging to obtain highly pure samples in sufficient amounts for crystal growth. This might change in the future with the recently reported use of porous complexes, which requires no crystallization of the target molecules (*17*).

The pair, 26-deoxyactein (**22**) and 23-*epi*-26-deoxyactein (**29**), serve as an excellent examples of the few *Actaea* triterpenes that have been characterized this way (*18*). X-ray diffraction analyses of these two congeners showed that the torsion angles around atom C-23 are different, revealing the difference in stereochemistry at that chiral center (see Fig. 5), which confirmed the structures and absolute configurations proposed by NMR for **22** and **29**. This result clarifies the structure of the compound previously referred to as 27-deoxyactein, as 23-*epi*-23-deoxyactein (**29**).
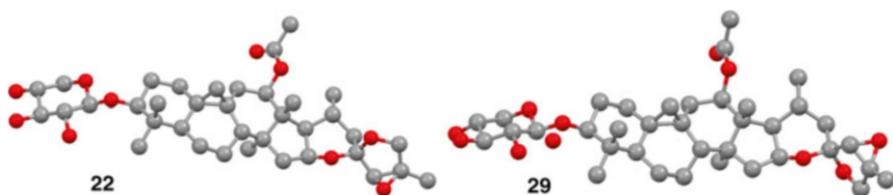
Fig. 5 The 3D structures of **22** and **29** without hydrogens drawn from original X-ray single crystal data (Cambridge Crystallographic Data Center as CCDC 170702 and 170703, respectively) with Mercury (ver. 3.1). The extra oxygen in the vicinity of C-1′ in 23-*epi*-26-deoxyactein is from a molecule of water

2.1.5 Dereplication

Dereplication is the process of differentiating and identifying one or more natural products present in extracts, fractions, or as isolated materials. Dereplication facilitates rapid identification and quantification of known compounds of interest, but also the identification of unknowns, from complex natural product mixtures. The dereplication of triterpenes from *A. racemosa* mainly utilizes two techniques: hyphenated LC (LC/ELSD, LC/MS) and NMR. The specifics of these dereplication approaches are described in the following.

2.1.5.1 Dereplication by LC/MS and LC/ELSD

LC-based methods are the most common for the dereplication of *Actaea* triterpenes. As *Actaea* triterpenes most commonly lack UV chromophores, UV spectroscopy is an inappropriate method for their LC detection. Instead, MS and ELSD are commonly used. For example, *Kan et al.* (*19*) developed an LC/(+)APCIMS method for the direct analysis and identification of four triterpenes, actein (**24**), 26-deoxyactein (**22**), cimiracemoside C (**2**), and cimigenoside (**1**), in *A. racemosa* and several commercially available *A. racemosa* products. An HPLC/MS total ion chromatogram of the triterpene-enriched sample was initially obtained, and identification of each triterpene was based on the presence of the molecular ion and characteristic fragment ions. This is well displayed in Figs. 1 and 2 of their publication (*19*).

Recently, *Avula et al.* used UPLC/ELSD for the identification and quantification of three triterpenes, actein (**24**), 23-*epi*-26-deoxyactein (**29**), and cimiracemoside F (**38**), in *A. racemosa* and dietary supplements derived from it (*20*). The marker compounds were readily identified by comparison of the UHPLC/ELSD chromatograms of plant samples and reference materials. This method was capable of giving shorter retention times while maintaining good resolution compared with conventional HPLC. In addition, UHPLC/MS was used for the identification of these three marker compounds. This method involved the use of $[M+Na]^+$ ions in the ESI mode for **24**, **29**, and **38** at retention times of 3.05, 5.10, and 5.40 min, respectively, and corresponding $m/z = 699.4$, 683.4, and 699.4 in the positive-ion mode with selected-ion monitoring (SIM).

2.1.5.2 Dereplication by NMR

While LC-based methods are used frequently in the dereplication of natural products, they provide limited information for structural identification, especially if performed with nominal mass resolution. The challenge increases further when the analytes exhibit weak UV absorption and/or poor MS ionization, such as the *Actaea* triterpenes. Moreover, LC-based methods rely on the availability of authentic reference materials for identification and particularly quantification. In contrast, NMR spectroscopy can serve as a more universal and linear detector for biologically complex

samples, and, thus, is a highly suitable tool for the dereplication of natural products. Not only does NMR typically provide structural information, but it allows the simultaneous quantification of each detected analyte in a complex matrix, frequently without the need for further separation, and more importantly, without identical calibrants (*16*, *21–23*). Despite these advantages, the complexity of the NMR spectra, such as peak multiplicity and signal overlap, leads to difficulties in the analysis of spectra.

One major challenge with dereplication of the *Actaea* triterpenes is to determine their structural subtypes, *i.e.* the partial structures of the aglycone side chains. From a general point of view, in the NMR spectra, any single chemical entity is represented by a unique pattern of NMR signals. Much like fingerprints, a sub-portion of these complex spectroscopic patterns might be sufficient to distinguish the different chemical entities. For *Actaea* triterpenes, the protons at C-16, C-22, C-23, and C-24 are excellent structural indicators, because they are all elements of the aglycone side chains and have distinct chemical shifts, multiplicities, and coupling constants (Table 10). In addition, their ^{1}H NMR signals are less overlapped in the region of 3.5–6.0 ppm, making them more easily recognized fingerprints for the identification of the structural subtypes of *Actaea* triterpenes.

In addition, the skeletal methyl groups of *Actaea* triterpenes can serve as powerful structural reporter groups. As shown in Table 11, each type of *Actaea* triterpene has a particular pattern of methyl signals in the ^{1}H NMR spectra, which encodes the structural characteristics of the triterpene skeletons. Therefore, as the Me ^{1}H signals are almost all singlets (except that of CH$_3$-21), they produce patterns that can be used for rapid dereplication in a manner similar to barcodes. Based on this hypothesis, *Qiu et al.* developed a robust and automatic dereplication system using only the methyl ^{1}H NMR signals for a group of *Actaea* triterpenes (*8*).

The concept of using methyl ^{1}H NMR shifts for structural dereplication of the *Actaea* triterpenes has a particular advantage in that methyl resonances are usually singlets of relatively high intensity. In approximation, comparing a ddd methylene (1H) with a singlet methyl (3H) signal, the individual spectral lines of the former are ~24-fold lower in intensity. Accordingly, minor triterpenes with a content of more than 4% become visible by virtue of their methyl signals even in crowded regions of the spectra.

Two statistical methods have been applied to establish the mathematical relationships between the methyl shifts and the structural characteristics of the *Actaea* triterpenes. In order to create a more universal dereplication system, an in-house NMR database was initially constructed to include not only the triterpenes identified in *A. racemosa*, but also those from many other *Actaea* species. Using a canonical discriminant analysis (CDA), all the *Actaea* triterpenes included in the in-house database could be classified with an overall correct rate of 86.9% (see Fig. 6).

In addition, Classification and Regression Tree Analysis (CART) was used to build a dereplication system for the *Actaea* triterpenes. CART is a machine-learning technique ideal for large and unbalanced data sets with many descriptors. It generates a tree-like graph or model as a binary-decision support tool to identify the origin or class of the samples being considered. Figure 7 shows the classification

Table 10 Characteristic ^{1}H NMR fingerprinting signals of the aglycone side chains of the five major classes of *Actaea* triterpenes

Compound type	Structure	Fingerprinting signal[a]	δ_H, m (*J*/Hz)
Acta-16,23;23,26-binoxol (Acteol)			4.630, q (7.3)
Acta-16,23;16,24-binoxol (Cimigenol)			4.780, d (9.0)
Acta-16,23;22,25-binoxol (Cimiracemoside)			3.920, d (11.0)
(24*R*)-24-Acetoxy-(16*R*)-16-hydroxyacta-16,23-monoxol (Hydroshengmanol)			5.650, d (8.0)
(23*R*)-23-Acetoxy-(24*S*)-24,25-epoxyactanol (23-*O*-Acetylshengmanol)			5.410, ddd (11.0, 8.0, 2.5)
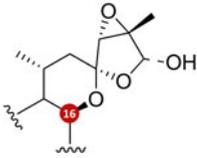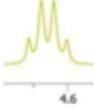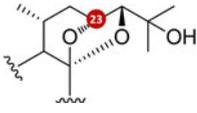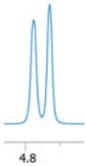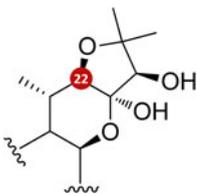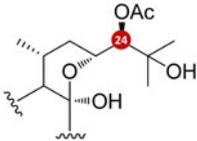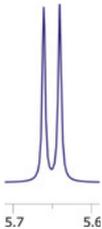
The multiplicity designations are made under first-order assumptions. Notably, the signals show nuances of higher order effects (*e.g.* roofing), which can be used in addition to their "basic multiplicity" for their assignment. As variations of the aglycone structure within each series will be mostly occurring in the A/B/C-rings, substituent chemical shift (s.c.s.) effects can be expected to be remote and will likely have no or very little effect on the shapes of these fingerprint signals
[a]The spectra were measured using pyridine-d_5 as the solvent at 400 MHz

binary tree (CBT) generated from CART analysis for the classification of *Actaea* triterpenes with seven skeletal methyl groups. This CBT is characterized by 13 terminal nodes and 12 non-terminal nodes, with an overall success rate of 94.4%. Leave-one-out cross-validation (LOOCV) indicates that the model has an excellent

Table 11 ^{1}H NMR "barcoding" of the skeletal methyl groups of the five major classes of *Actaea* triterpenes

Compound type	Structure	^{1}H NMR "barcode" of skeletal methyls[a]
Acta-16,23; 23,26-binoxol (Acteol)		
Acta-16,23; 16,24-binoxol (Cimigenol)		
Acta-16,23; 22,25-binoxol (Cimiracemoside)		
(24*R*)-24-Acetoxy-(16*R*)-16-hydroxyacta-16,23-monoxol (Hydroshengmanol)		
(23*R*)-23-Acetoxy-(24*S*)-24,25-epoxyactanol (23-*O*-Acetylshengmanol)		

[a]Each spectrum was constructed based on the average values of the methyl shifts of the same type of *Actaea* triterpenes. The spectra only illustrate the distribution pattern of the methyl signals, not the intensity and line shape of the signals, which are irrelevant for this purpose

prediction rate for the majority of the *Actaea* triterpenes (cimigenols 80.0%, cimiracemosides 91.7%, and hydroshengmanols 100%). This means the CBT model can not only be used for the dereplication of known compounds, but also it has the potential to identify the aglycone type of compounds yet to be discovered. Considering inescapable variations of reported ^{1}H chemical shift information due to inconsistencies in, *e.g.* temperature and calibration (TMS *vs.* residual solvent), the discriminative power of the model could be improved further by future use of a standardized NMR acquisition protocol, such as 1D qHNMR protocols (*22*, *23*).
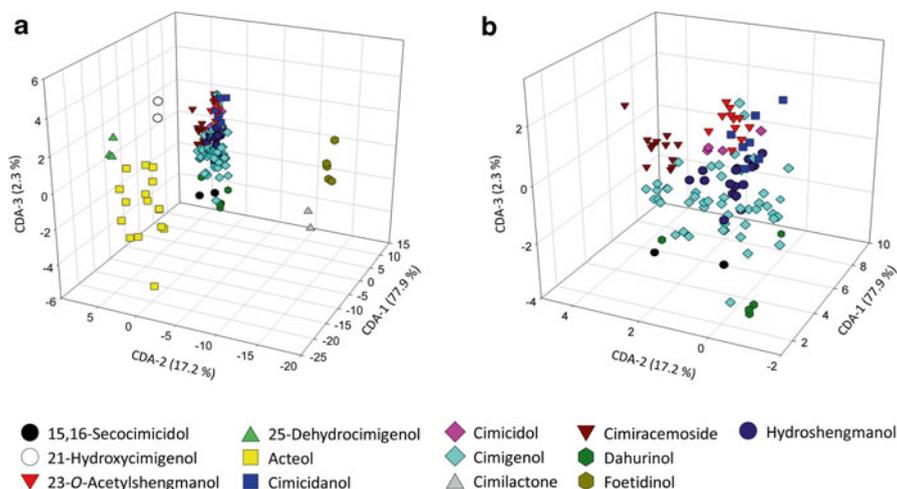
Fig. 6 Panel **a** shows that the first three factors account for 97.4% of the total variance in the Me shifts of the *Actaea* triterpenes. Panel **b** shows the sub-cluster of all triterpenes with seven skeletal Me groups having CDA-1 scores between 0 and 10, which form further sub-clusters depending on the specific skeleton types [taken from *Qiu et al.* (8) with permission from the American Chemical Society]
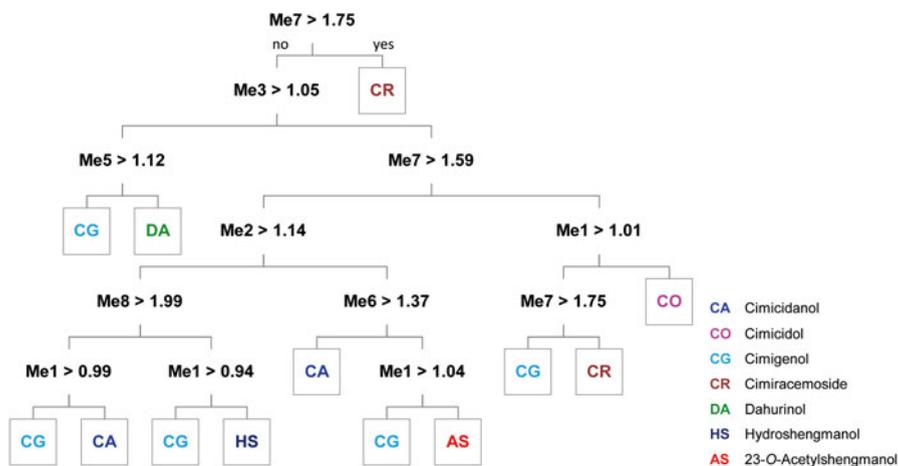
Fig. 7 The CBT model for classification of *Actaea* triterpenes with seven skeletal Me groups by using the Me ^{1}H NMR shifts as descriptors (Me1>Me2>…>Me8). Methyls are numbered according to their chemical shift from the most upfield one (8)

The remaining structural elements, such as substituents and sugar moieties, are more readily recognized by their fingerprinting signals in the ^{1}H NMR spectra. For example, the occurrence of the double bond at C-7 can be deduced from the ^{1}H chemical shifts of cycloartane protons, *e.g.* H-19a, a doublet (d) signal at ~1.00 ppm. The sugar moieties, commonly xylose and arabinose, can be differentiated by the multiplicity pattern of their H-5′b signals: a triplet (t) at ~3.75 ppm with $J = \sim 11$ Hz for xylose, and a doublet of doublets (dd) at ~3.80 ppm with J ~13 and ~2 Hz for arabinose. The acetoxy substituents at C-12 can be readily identified by the singlet methyl signal at ~2.10 ppm, and confirmed by observing a dd signal ($J \sim 9$ and ~4 Hz) for H-12 at ~5.10 ppm.

2.2 Cimicifugic Acids

2.2.1 Structural Characteristics

Cimicifugic acids (Table 12) are condensation products of a benzyltartaric acid moiety, such as fukiic acid, and a cinnamoyl moiety, as exemplified by caffeic acid. Several cimicifugic acids were isolated from the aqueous ethanol extract of BC, including cimicifugic acid KC (fukinolic acid (**53**)), cimicifugic acid KF (cimicifugic acid A, (**54**)), cimicifugic acid KI (cimicifugic acid B, (**55**)), cimicifugic acid PF (cimicifugic acid E, (**56**)), cimicifugic acid PI (cimicifugic acid F, (**57**)), and cimicifugic acid KS (**58**). Furthermore, the acyl residues of **53–58**, *i.e.* caffeic (**48**), ferulic (**49**), isoferulic (**50**), piscidic (**51**), and fukiic acids (**52**), were isolated as the free acids, also.

2.2.2 Naming System

Since the initial discovery of fukinolic acid, historical nomenclature has evolved by giving fukinolic acid congeners isolated from *Cimicifuga/Actaea* species the names of cimicifugic acids, adding alphabetical letters in the order of their discovery. Structurally, cimicifugic acids are dimeric molecules, consisting of a benzyl tartaric acid moiety and a cinnamoyl moiety, either of which can exhibit variations in the substitution pattern of their respective aromatic rings. In order to directly connect their structural features with the compound names, a new nomenclatural system was established for cimicifugic acids (*24*). In this systematic nomenclature, two letters are added to the name of the compound class, cimicifugic acid, to indicate the substitution pattern of the two aromatic rings, Ar_A and Ar_B (see Fig. 8). Each letter originates from the abbreviation of the trivial name of the similarly substituted phenylpropanoic moiety (see Sect. 2.2.1).

Table 12 Structures of (**A**) the phenylpropanoic acids and their derivatives, including the cimicifugic acids (**B**)

A	R^1	R^2	Common name
48	OH	OH	Caffeic acid (C)
49	OCH_3	OH	Ferulic acid (F)
50	OH	OCH_3	Isoferulic acid (I)

	R	Common name
51	H	Piscidic acid (P)
52	OH	Fukiic acid (K)

B	R^1	R^2	R^3	R^4	Common name	New nomenclature
53	OH	OH	H	OH	Fukinolic acid	Cimicifugic acid KC
54	OCH_3	OH	H	OH	Cimicifugic acid A	Cimicifugic acid KF
55	OH	OCH_3	H	OH	Cimicifugic acid B	Cimicifugic acid KI
56	OCH_3	OH	H	H	Cimicifugic acid E	Cimicifugic acid PF
57	OH	OCH_3	H	H	Cimicifugic acid F	Cimicifugic acid PI
58	OCH_3	OH	OCH_3	OH	–	Cimicifugic acid KS
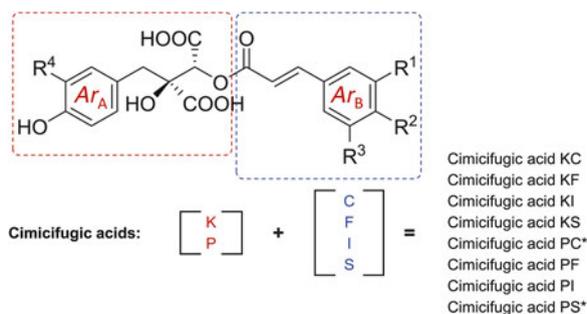
Fig. 8 The rational naming system for cimicifugic acids. The compounds denoted by asterisks are yet to be found in *A. racemosa*

2.2.3   Isolation Techniques

Cimicifugic acids are phenolic compounds, which bear several acidic hydroxy groups. Therefore, polyamide column chromatography is an excellent technique for their separation. The selective interaction with the polyamide stationary phase results from the formation of intermolecular hydrogen bonding between the carboxyl groups of the polyamide and the hydroxy groups of the analytes. Common protocols include the use of an aqueous ethanol BCE and the performance of an initial fractionation using polyamide column chromatography, with aqueous methanol for elution. This affords an enriched fraction of cimicifugic acids. Further separation of individual compounds is then performed on repeated RP-MPLC (*e.g.* C_{18}). Preparative RP-HPLC has usually been employed in the last purification step to resolve the congeners that are otherwise difficult to separate by RP-MPLC (*25*).

   Recently, a protocol for pH-zone refining centrifugal partition chromatography (pHZR-CPC) has been established to separate cimicifugic acids (*24*). This methodology was developed based on the hypothesis that the cimicifugic acids occur in the BCE in association with bases (alkaloids), some of which are responsible for the observed 5-HT_7 activity. The separation procedures started with ethyl acetate–water partitioning of the methanolic BCE. The aqueous layer was then subjected to column chromatography on Amberlite XAD-2, which was eluted sequentially with water and methanol. The methanolic eluate was further subjected to pHZR-CPC in the RP mode using the solvent system ethyl acetate–butanol–water (1:4:5). The pH was adjusted by adding 0.3% (*v/v*) of a 28% aqueous ammonia solution and 0.1% (*v/v*) of trifluoroacetic acid (TFA) in the mobile (lower) and stationary (upper) phases, respectively. This facilitated the separation of the acids from their basic complex-partners. Finally, the enriched fraction of the dissociated acids was separated in a second step of CPC with the solvent system HEMWat +5 (*n*-hexane–ethyl acetate–methanol–water, 3:7:3:7, with 0.2% of TFA in both phases) to afford the pure compounds.

2.2.4   Structural Elucidation

Combined use of mass spectrometry and NMR spectroscopy is an efficient approach for the structural elucidation of cimicifugic acids. Initially, ESI-MS can be used for the determination of the molecular weight of the compounds. Typically, negative-mode ESI is employed because it has been found to be more sensitive for these phenolic acids. Further structural elucidation is typically aided by 1D- and 2D-NMR methods, including ^{1}H, ^{13}C, COSY, HSQC, and HMBC experiments (*24*, *25*).

2.2.5   Dereplication

One-dimensional ^{1}H-NMR spectra can serve as highly characteristic fingerprints for the cimicifugic acids. *Gödecke et al.* (*24*) reported the use of a structure-based spin-pattern analysis for the dereplication of known cimicifugic acids. This method is based on the ^{1}H NMR spectroscopic differences resulting from the characteristic substitution patterns on the aromatic rings of the acids (see Fig. 9). For example, the Ar_A rings of monooxygenated acids give rise to an AA′XX′ spin resonance that is distinctly different from the AMX pattern of their 3,4-dioxygenated counterparts. The chemical shifts of the signals of the Ar_B ring are affected by the position of methoxylation, as opposed to hydroxylation. In addition, the signals of the methoxy protons also exhibit shift variations dependent on their position in the Ar_B ring.

2.3   Nitrogen-Containing Constituents

The first reported occurrence of alkaloids in *A. racemosa* was reported at the beginning of the last century. *Horace Finnemore* published in the *Pharmaceutical Journal* in 1909 (*26*) that he had detected trace amounts of alkaloids in both basic and acidic extracts of BC, although he did not indicate what kind of alkaloid reagent he used. There was no further report of alkaloids from *A. racemosa* until 1956, when *Gemeinhardt et al.* reported the finding of *N*-methylcytisine (**59**) from *A. racemosa* (*27*, *28*).
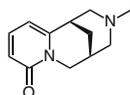
59 (*N*-methylcytisine)

However, this type of alkaloid has been found usually in the genera *Cytisus* (Fabaceae) and *Caulophyllum* (Berberidaceae), which are generally prolific in alkaloids. Cytisine was first isolated from *Cytisus laburnum* in 1862 (see (*29*) and references therein). However, from a biosynthesis perspective, it seems unlikely that *N*-methylcytisine was from *Actaea*. It is more likely that the plant material in *Gemeinhardt*'s research was misidentified.
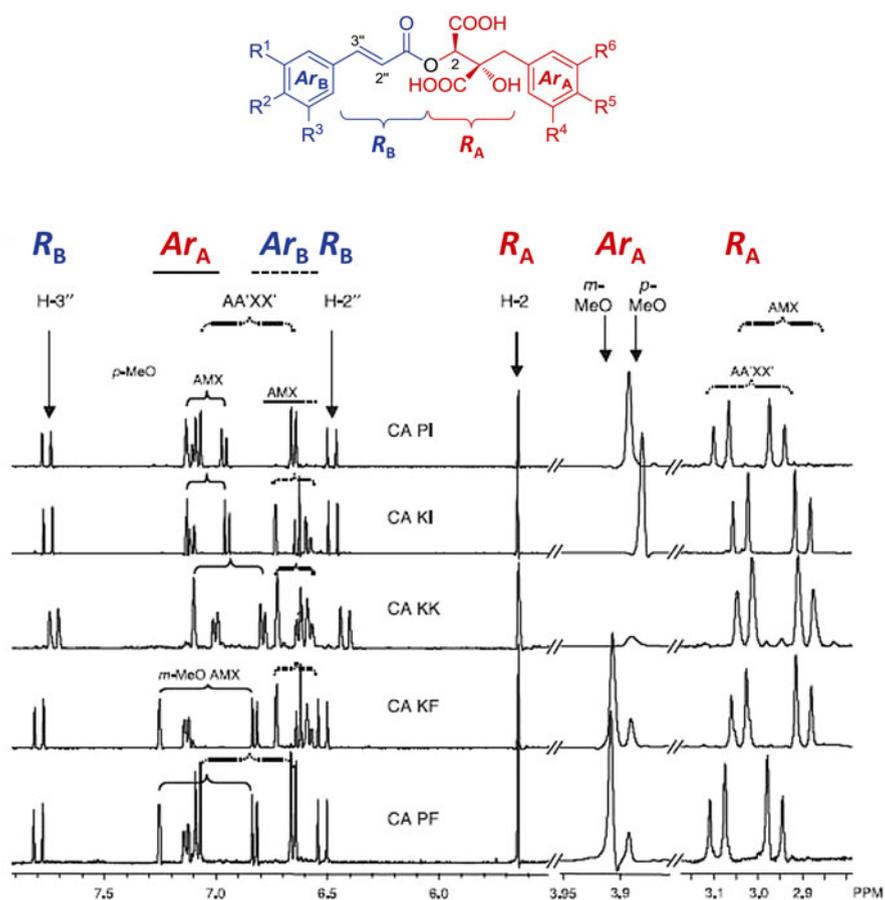
Fig. 9 ^{1}H NMR spin pattern of cimicifugic acids: KF, KI, KK, PF, and PI (400 MHz, CD$_3$OD, chemical shift reference set to 3.21 ppm for residual methanol-d_3, relative to TMS signal at 0.0 ppm as external standard). According to (*24*), reprinted with permission from Wiley Ltd

In 1965, *Crum et al.* screened alkaloids in different parts of seven native Ohio plants (*30*). The authors used basic and acidic approaches to prepare the extracts. Five different alkaloid reagents were used to detect alkaloids: (1) *Dragendorff* reagent: bismuth potassium iodide; (2) *Mayer* reagent: mercuric potassium iodide; (3) Silicotungstic acid reagent: 12% aqueous silicotungstic acid; (4) *Sonnenschein* reagent: phosphomolybdic acid; and (5) *Wagner* reagent: iodine-potassium iodide. The criterion was to determine if the investigated plant or plant part tested positive for alkaloids was that the reaction "gave, as a minimum, a distinct turbidity with four of the five reagents in either the acid or base extraction method". The use of five different reagents eliminated the commonly observed "false-positive" results when only using one reagent. For

instance, the most widely used alkaloid reagent, *Dragendorff* reagent, gave a false-positive color reaction when 2-pyrones were present in the extract (*31*). Based on *Crum et al.*'s screening procedure, no alkaloids were detected in acidic and basic extracts of leaves and stems or roots/rhizomes, but occurred in an acidic extract of the seeds. While the seeds of *A. racemosa* have not been used as a drug or for medicinal purposes, this finding was the first indication of the occurrence of alkaloids in *A. racemosa*.
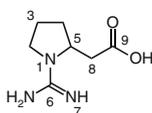
60 (cimipronidine)

In 2005, *Fabricant et al.* were the first to report a new type of guanidine alkaloid, cimipronidine (**60**), as a major constituent from an *n*-BuOH soluble fraction of a methanolic BCE (*32*). This finding clarified the zwitterionic nature of some of the *Actaea* alkaloids and subsequently changed the phytochemical protocols used in the same laboratory to further investigate the plant.
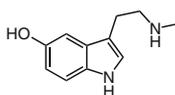
61 (N_ω-methylserotonin)

Successively more alkaloids were isolated and/or identified from BC. In 2008, the serotonergic active principle, N_ω-methylserotonin (**61**) (*33*), was identified from the highly retained fraction of a pH-zone refining fast centrifugal partition chromatography (FCPC) procedure separating the tertiary subfraction from a methanolic crude extract. In addition to two cimipronidine derivatives (**62** and **63**), *Gödecke et al.* isolated and identified salsolinol (**64**) and dopargine (**65**) (*34*), which represent another class of alkaloids in BC. These alkaloids can be explained as representing products of dopamine with an aldehyde (acetaldehyde or γ-guanidinobutraldehyde) as substrates through a *Pictet-Spengler* reaction with an enzyme, such as strictosidine or norcoclaurine synthase.
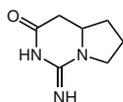
62 (cyclocimipronidine)   **63** (cimipronidine methyl ester)   **64** (salsolinol)   **65** dopargine
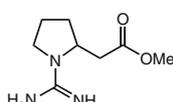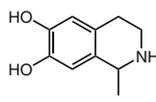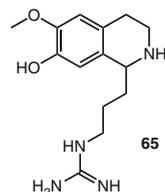
In 2012, *Nikolić et al.* published a detailed high-resolution LC-MS analysis of the nitrogen-containing components in a standardized 75% EtOH BCE. The extract was the same material that was used in Phase I and Phase II-b clinical trials conducted by the University of Illinois at Chicago (UIC)/National Institutes of Health

(NIH) Botanical Center (*35*). Using MS^2 accurate mass measurements, a combination of database searches, known biosynthesis pathways of alkaloids, and comparison with authentic standards and their fragmentation patterns, the authors identified 73 nitrogen-containing compounds with varying degrees of confidence. These results revealed that BC contains a large number of remarkably diverse and previously unreported nitrogen-containing metabolites. The observed compounds covered both primary and secondary metabolites and included derivatives of amino acids, nucleobases, aporphines, betaines, β-carbolines, cholines, cinnamides, guanidines, isoquinolines, and protoberberines (Fig. 10).
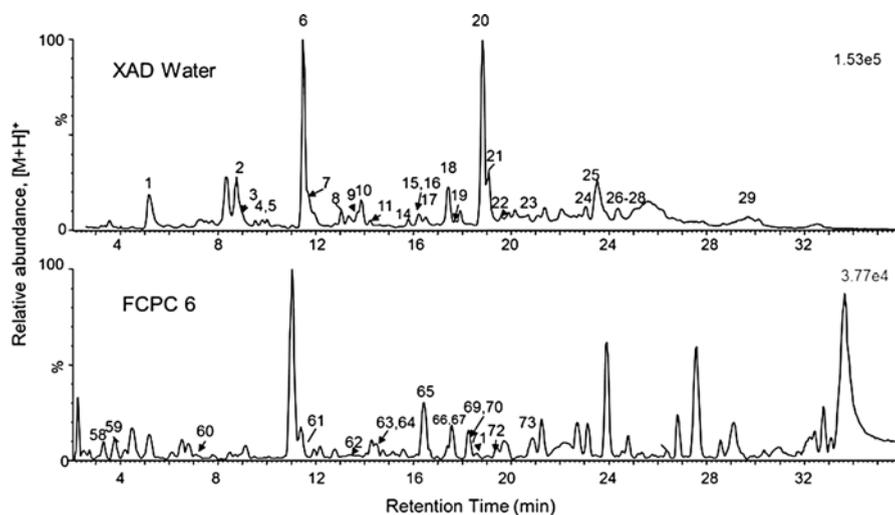
Fig. 10 XAD-2 water fraction (*top*) and the last fraction (*bottom*) of pH-zone refining FCPC separation of the MeOH fraction of XAD-2 chromatography from the extract (from *Nikolić et al.* (*35*) with permission of the Elsevier B.V.) The compounds discovered in this study were, for the most part, characterized by mass spectrometry. Hence no stereochemistry was shown, even in cases where the stereochemistry could be assumed with reasonable confidence

2.3.1   Structural Diversity

2.3.1.1   Primary Metabolites

Amino Acid Derivatives

This group consists of many amino acids. As well as two nonessential amino acids, pipecolic (**66**) and pyroglutamic (**67**) acid, others are derivatives of essential amino acids with a feruloyl or isoferuloyl group attached to the α-amino group. Eight arginine (**68–75**), two phenylalanine (**75, 76**) and one histidine (**78**) derivatives were identified in a BCE.

66 (pipecolic acid)   **67** (pyroglutamic acid)

76 R = H (phenylalanine)
77 R = feruloyl-4-*O*-hexoside
(α-*N*-feruloylphenylalanine-4-*O*-hexoside)

68 R^1 = R^2 = H (arginine)
69 R^1 = formyl, R^2 = H (α-*N*-formyl arginine)
70 R^1 = acetyl, R^2 = H (α-*N*-acetyl arginine)
71 R^1 = caffeoyl, R^2 = H (α-*N*-caffeoyl arginine)
72 R^1 = feruloyl, R^2 = H (α-*N*-feuroyl arginine)
73 R^1 = isoferuloyl, R^2 = H (α-*N*-Isofeuroyl arginine)
74 R^1 = isoferuloyl, R^2 = CH$_3$ (α-*N*-isofeuroyl arginine methyl ester)
75 R^1 = isoferuloyl, R^2 = CH$_2$CH$_3$ (α-*N*-isofeuroyl arginine ethyl ester)

78 (α-*N*-isoferuloyl histidine)

Nucleobase Derivatives

As basic units of genetics, nucleobases are found within DNA, RNA, nucleotides, and nucleosides in all organisms. In this group, four adenine derivatives (**79–82**) were identified from a 75% alcoholic BCE. The two other nucleobases found were cytidine (**83**) and guanosine (**84**).

79 R^1 = R^2 = H (adenine)
80 R^1 = ribofuranoside, R^2 = H (adenosine)
81 R^1 = ribofuranoside, R^2 = CH$_3$ (6-methylaminoadenosine)
82 R^1 = 2'-methylribofuranoside, R^2 = H (2'-methyladenosine)

83 (cytidine)   **84** (guanosine)
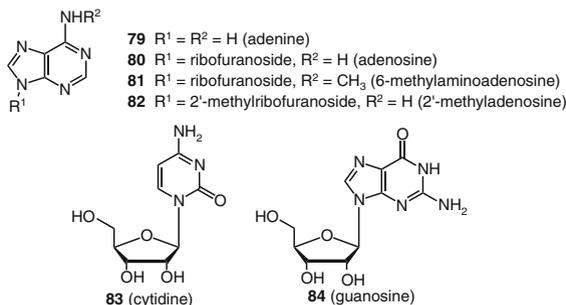
2.3.1.2   Secondary Metabolites

Aporphines

Aporphines are a class of isoquinoline alkaloids. More than 500 aporphine alkaloids have been isolated from various plant families, and many of these compounds display potent cytotoxic activities, which may be exploited for the design of anticancer agents (*36*). From this group, six known compounds (**85–90**) were identified from BC.
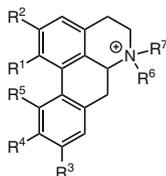
85 R^1 = OH, R^2 = OCH_3, R^3 = H, R^4 = OCH_3, R^5 = OH, R^6 = R^7 = CH_3 (magnoflorine)

86 R^1 = OH, R^2 = OCH_3, R^3 = OH, R^4 = OCH_3, R^5 = H, R^6 = R^7 = CH_3 (laurifoline)

87 R^1 = OCH_3, R^2 = R^3 = H, R^4 = OCH_3, R^5 = R^6 = R^7 = H (laurolitsine)

88 R^1 = R^2 = OCH_3, R^3 = H, R^4 = OCH_3, R^5 = OH, R^6 = R^7 = CH^3 (menisperine)

89 R^1 = R^2 = OCH_3, R^3 = OH, R^4 = OCH_3, R^5 = R^6 = R^7 = H (lauriotetanine)

90 R^1 = R^2 = OCH_3, R^3 = OH, R^4 = OCH_3, R^5 = H, R^6 = R^7 = CH^3 (xanthoplanine)

Betaines

As compatible solutes or osmoprotectants, betaines are an important class of naturally occurring compounds to balance the osmotic difference between the cell's surroundings and the cytosol. Five BC metabolites were identified as belonging to this group (**91–95**) (*37*).

91 (carnitine)   **92** (glycine betaine)   **93** (histidine betaine)   **94** (proline betaine)   **95** (trigonelline)
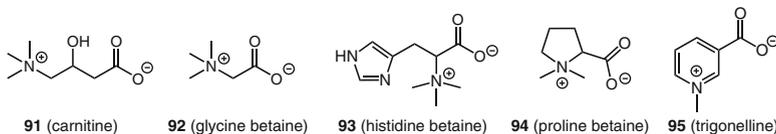
β-Carbolines

β-Carboline alkaloids are one structural sub-type of the indole alkaloids. Previously, most research on the β-carbolines alkaloids focused on effects on the central nervous system, such as being inverse agonists of benzodiazepine receptors, and having an affinity for 5-HT_{2A} and 5-HT_{2C} receptors (*38–40*). Recently, research has revealed many new pharmacological effects relevant to antitumor (*41*), antiviral (*42*), antimicrobial (*43*), and antiparasitic activities (*44*). Four of these alkaloids (**96–99**) were identified by LC-MS/MS, and the structure of cimitrypazepine (**96**) was confirmed by synthesis.

Cinnamides

Amides of ferulic and isoferulic acids were detected in a BCE. This class of compounds has previously shown broad biological properties, including central nervous system depressant, anticonvulsant, muscle relaxant, antiallergic, antineoplastic, and anti-infective activities. The three cinnamides (**100–102**) were identified in a BCE.

96 (cimitrypazepine)   **97** (1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid)

98 (N(2)-methyl-6-hydroxy-3,4-dihydro-β-carboline)   **99** (N(2)-methyl-6-hydroxy-1,2,3,4-tetrahydro-β-carboline)

100 R = feruloyl (feruloyl putrescine)
101 R = isoferuloyl (isoeruloyl putrescine)

102 (N-isoferuloyl glutamic acid)
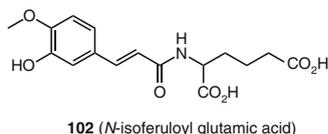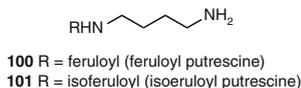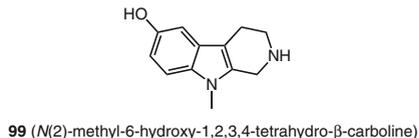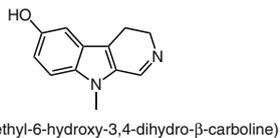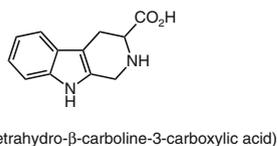
Cholines

Cholines generally refer to the various quaternary ammonium salts that contain the partial structures of derivatives of N,N,N-trimethylethanol-ammonium cation. Choline itself is a water-soluble essential nutrient. Egg and fatty meats are high choline-containing foods. Despite the perceived benefits of cholines, dietary recommendations have discouraged eating high-choline foods. However, in 2005, a National Health and Nutrition Examination Survey stated that only 2% of post-menopausal women consume the recommended intake for choline. So far, a total of five cholines (**103–107**) have been identified from a BCE.
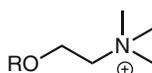
103 R = H (choline)
104 R = benzoyl (benzoyl choline)
105 R = feruloyl (feruloyl choline)
106 R = isoferuloyl (isoferuloyl choline)
107 R = hexosyl (hexosyl choline)

Guanidines

Besides the above-mentioned three guanidines, **60**, **62** and **63**, a related structure to this type of alkaloid, **108**, was identified using LC-MS. It has a similar fragment pattern to that of cyclocimipronidine (**62**) and cimipronidine methyl ester (**63**), except for having 14 Da more and 18 Da less in its molecule ion than do **62** and **63**, respectively, which led to the assignment of the methyl group to the amino group at the C-7 position, forming a methylamine group.
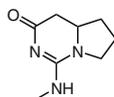
108 (*N*-methylcyclocimipronidine)

Isoquinolines

Since the first isoquinoline derivative was isolated from coal tar in 1885, more than 400 simple isoquinoline derivatives have been reported. This class of alkaloids shows many different bioactivities, from antimalarial, anti-HIV, antineoplastic, antimicrobial, and antibacterial to insect growth retardation. Several tetrahydroisoquinoline derivatives have been found to act as neurotoxin precursors linked to *Parkinson*'s disease. In addition to dopargine (**63**), six hydrogenated isoquinoline alkaloids, including (**109–110**) and benzyl substituted (**111–114**) derivatives, were found in BCE.
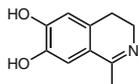
109 (1,2-dehydrosalsolinol)   **110** (norsalsolinol)
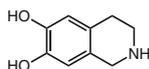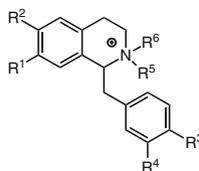
111 R^1 = OH, R^2 = OCH$_3$, R^3 = OH, R^4 = H, R^5 = R^6 = CH$_3$ (magnocurarine)
112 R^1 = R^2 = R^3 = OH, R^4 = R^5 = R^6 = H (norcoclaurine)
113 R^1 = OH, R^2 = R^3 = OCH$_3$, R^4 = OH, R^5 = H, R^6 = CH$_3$ (reticuline)
114 R^1 = OCH$_3$, R^2 = R^3 = OH, R^4 = H, R^5 = R^6 = CH$_3$ (oblongine)

Protoberberines

The protoberberine alkaloids are derived biogenetically from tyrosine. The quaternary protoberberine alkaloids represent approximately 25% of all currently known alkaloids with a protoberberine skeleton isolated from natural sources. The protoberberines are mainly distributed in the Papaveraceae, Berberidaceae, Fumariaceae, Menispermaceae, Ranunculaceae, Rutaceae, and Annonaceae. More than 100 berberine alkaloids, including tetrahydroprotoberberines, and quaternary protoberberines were reported. The bioactivities of protoberberines (*45*, *46*) include the inhibition of DNA synthesis, protein biosynthesis, membrane permeability, and the uncoupling of oxidative phosphorylation. From a BCE, two protoberberine alkaloids were identified tentatively by LC-MS/MS, but their unambiguous identification could not be achieved due to the lack of authentic standards. As a result, the structures of phellodendrine (**115a**) or cyclanoline (**115b**), and *N*-methyltetrahydrocolumbamine or its isomer (**116**) could be proposed by comparison of their fragmentation patterns with those of structural analogs.
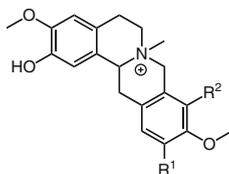
115a R^1 = OH, R^2 = H (phellodendrine)
115b R^1 = H, R^2 = OH (cyclanoline)
116   R^1 = H, R^2 = OCH$_3$ or isomer (*N*-methyl tetrahydrocolumbamine)

Protopines

Protopine is a benzylisoquinoline alkaloid occurring in the genera *Corydalis* and *Fumaria* belonging to the family Papaveraceae. The title compound has been found to inhibit histamine H1 receptors and platelet aggregation, and can act as an analgesic. Two protopine alkaloids (**117**, **118**) were identified on the basis of MassBank database searches and were confirmed by comparison with authentic standards (*47*).
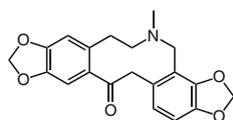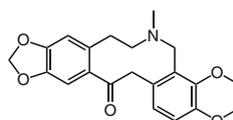
117 protopine   **118** allocryptopine

Pyridoxines

Vitamin B$_6$ is the classic pyridoxine, and analogs of this simple compound have been described as helpful to balance hormonal changes in women and aid the immune system. Two compounds belonging to the pyridoxines (**119**, **120**) were identified on the basis of MassBank database searches and confirmed by comparison with authentic standards.
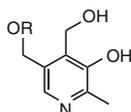
119 R = H (pyridoxine)
120 R = *β*-D-glucopyranosyl-(5'-*O*-*β*-D-glucopyranosylpyridoxine)

Miscellaneous

Several other nitrogen-containing components were also identified from BCE including **121–134**. The guanidines **121–127** share the highly basic group with the cimipronidines (**60**, **62**, and **63**) as well as the salsolinol derivative **65**. A general observation is that the BC alkaloids include a variety of guanidines. Similarly, the dopamides

128–131 share certain structural similarities with cimicifugic acid (Sect. 2.2), and it is not likely that hybrid molecules are present in BC.
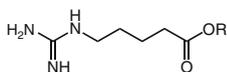
121 R = H (δ-guanidinovaleric acid)
122 R = CH$_3$ (methyl δ-guanidinovalerate)
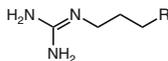
123 R = CH$_2$OH (γ-guanidinobutanol)
124 R = CHO (γ-guanidinobutyraldehyde)
125 R = CO$_2$H (γ-guanidinobutyric acid)
126 R = CO$_2$CH$_3$ (methyl γ-guanidinobutyrate)
127 R = CO$_2$CH$_2$CH$_3$ (ethyl γ-guanidinobutyrate)

132 (*N*-phenylacetyl acetamide)

128 R^1 = CH$_3$, R^2 = hexose, R^3 = R^4 = OH (*N*-feruloyl dopamine-4'-*O*-hexoside)
129 R^1 = hexose, R^2 = CH$_3$, R^3 = R^4 = OH (*N*-isoferuloyl dopamine-3'-*O*-hexoside)
130 R^1 = CH$_3$, R^2 = H, R^3 = H, R^4 = *O*-hexoside (*N*-feruloyl tyramine-4-*O*-hexoside)
131 R^1 = CH$_3$, R^2 = H, R^3 = OCH$_3$, R^4 = *O*-hexoside (*N*-feruloyl-3-*O*-methoxytyramine-4-*O*-hexoside)
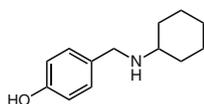
133 (*N*-cyclohexyl-4-hydroxybenzylamine)
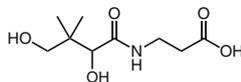
134 (pantothenic acid)

2.3.2 Structural Elucidation

Although alkaloids and other nitrogen-containing compounds are often present at relatively low abundance in plant extracts, this is offset by the fact that they may have potent bioactivities. Accordingly, small quantity changes can lead to large differences in the observed activities of crude extracts. For example, N_ω-methylserotonin (**61**) shows 5-HT$_7$ receptor binding activity ($IC_{50} = 23$ pM), induces cAMP ($EC_{50} = 22$ nM), and blocks serotonin re-uptake ($IC_{50} = 490$ nM) (*33*). Small variations of the absolute content of this marker compound in a BCE preparation would cause major changes in their biological potency and greatly affect biological standardization, while requiring high-sensitivity chemical analysis for quantitation. LC-MS/MS plays a crucial role in the identification of such minor components. High resolution LC-MS/MS using a Q-TOF instrument, analysis of characteristic fragmentation patterns, and database searches (*e.g.* MassBank) have been demonstrated to be a viable approach to identify a major portion of the alkaloids (~70%, based on LC-MS full-scan peak intensity) in a BCE with the highest confidence level 1, which is established by comparing the retention time and fragmentation pattern of an unknown with those of an authentic standard (*48*). Of the remaining substances from a total of 73 alkaloids, 26% were identified with confidence level 2, *i.e.* by comparing their tandem mass spectra either with published spectra or with tandem mass spectra of close structural analogs. Only 4% of the alkaloids described in this study could be identified at no better than confidence level 3. At this level, characterization can only ascertain the chemical class of an unknown based on the

similarity of tandem mass spectra with known compounds of the same class. A summary of general LC-MS/MS identification strategies and the characteristic fragmentation patterns of key alkaloids found in BC has been reported recently (*35*).

2.3.2.1 Nucleobases and Their Derivatives

Most nucleobases and nucleosides from BC were identified by spectral database searching and comparison with authentic standards. The shared elemental composition of methylated adenosine **81** and **82** ($C_{11}H_{15}N_5O_4$) was of particular interest. Besides a few product ions, the compounds had different product ion tandem mass spectra. The product ion tandem mass spectrum of **81** exhibited only one fragment ion (*m/z* 150, [MH − 133]$^+$), which is protonated after the loss of a sugar moiety. Its elemental composition corresponds to methylated adenine ($C_6H_8N_5$). Database searches revealed that methylation most likely occurred on the amino group, which led to tentative identification of this compound as *N*-methyladenosine (confidence level 3). In contrast, two fragment ions were observed in the tandem mass spectrum of **82**, one was protonated adenine (*m/z* 136 [MH − 147]$^+$) after loss of a sugar moiety, and another corresponded to loss of ammonia from adenine (*m/z* 119, [MH − 147 − 17]$^+$). Comparing its fragments with those of adenosine, this compound has an extra CH_2 unit in the sugar moiety. Based on literature searches, **82** is the most likely 2′-*O*-methyladenosine (confidence level 3).

2.3.2.2 Guanidino Alkaloids

Guanidino alkaloids, either acyclic or cyclic, produce a characteristic loss during collision-induced dissociation (CID) in the MS. For acyclic guanidino alkaloids or arginine derivatives (including **68–75**, **121–127**), the characteristic loss is a neutral guanidine (59 Da; CH_5N_3). To distinguish this neutral guanidine loss from other equal losses of 59 Da, an accurate mass measurement of the fragment is necessary. For example, acetamide (C_2H_5NO) is equal to guanidine with a MW of 59 Da. However, both can be distinguished readily by the elemental compositions. In addition to the loss of neutral guanidine, protonated guanidine [$(NH_2)_3C$]$^+$ of *m/z* 60 can usually be observed in the acyclic alkaloids. It has been proposed that guanidine is protonated *via* an ion-neutral complex (Scheme 1), and that its abundance is inversely proportional to the applied collision energy. Although the collision energy was applied as low as 5 eV, this ion was still not abundant, at only ~35% (*49*).
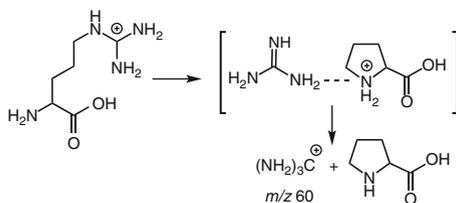
Scheme 1 Proposed fragmentation pathway of guanidino alkaloids

Representing one group of the BC guanidino alkaloids, the fragmentation pattern of the arginine derivatives, **69–75**, is very similar to that of arginine (**68**) itself (see Fig. 11). The presence of an ion at *m/z* 60 indicated that there was no substitution on the guanidino group, and that acetylation occurred on the amino group. Further, the identification of α-*N*-acetylarginine was confirmed by comparison with an authentic standard.
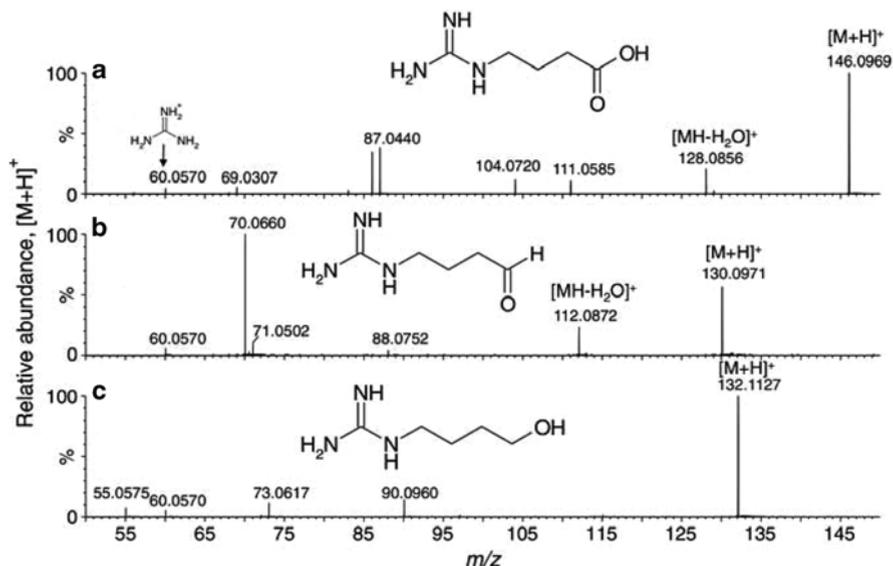
Fig. 11 Product ion tandem mass spectra of (**a**) arginine (**68**), (**b**) *N*-acetylarginine (**70**), and (**c**) *N*-formyl-arginine (**69**). From *Nikolić et al.* (*35*), with permission of Elsevier B.V.

For another group of guanidino alkaloids, γ-guanidinobutanol and its analogs (**122–127**), a characteristic neutral loss of 59 Da was observed at *m/z* 87 for **125**, *m/z* 71 for **124**, and *m/z* 73 for **123**, indicating that they all contain free guanidine groups (see Fig. 12).

The compound, γ-guanidinobutyric acid (GBA, **125**), was identified by database searches and confirmed with an authentic standard. The fragmentation pathway was proposed as shown in Scheme 2. The product ion of *m/z* 87 is a protonated butyro-lactone formed through an S_N2 attack of the carbonyl oxygen on the carbon atom bearing the guanidino group (see Scheme 2a). This is supported by observation of protonated guanidine at *m/z* 60. At higher collision energies, protonated guanidine is not observed due to insufficient survival time of the ion-neutral complex. An additional minor pathway for formation of the ion of *m/z* 87 is by elimination of ammonia from protonated γ-aminobutyric acid (GABA; *m/z* 104), as determined in separate ion-trap experiments. Similarly, the product ion of *m/z* 86 has an elemental composition of C_4H_8NO corresponding to protonated butyrolactam. Ion-trap experiments indicated that the main pathways of this ion are a loss of carbon diimine
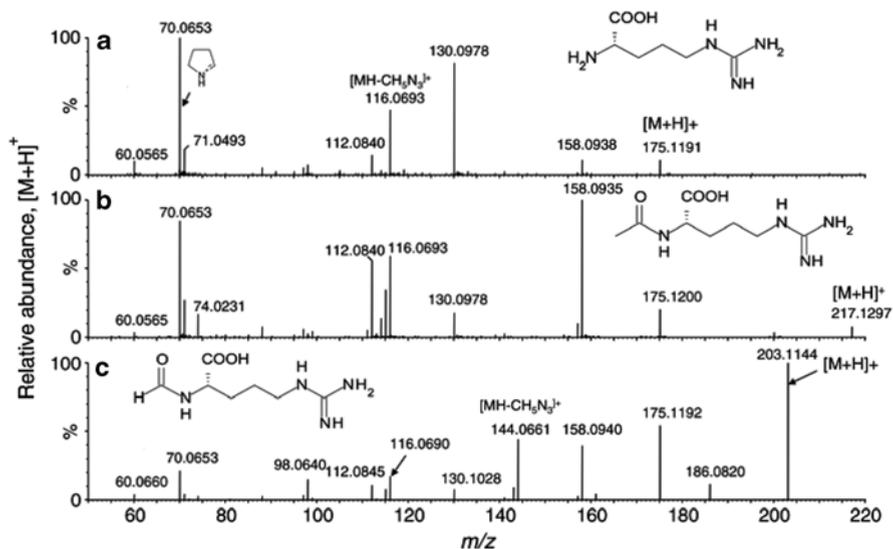
Fig. 12 Product ion tandem mass spectra of (**a**) γ-guanidinobutyric acid (**125**), (**b**) γ-guanidino-butyraldehyde (**124**), and (**c**) γ-guanidinobutanol (**123**). From *Nikolić et al.* (*35*), with permission of Elsevier B.V.
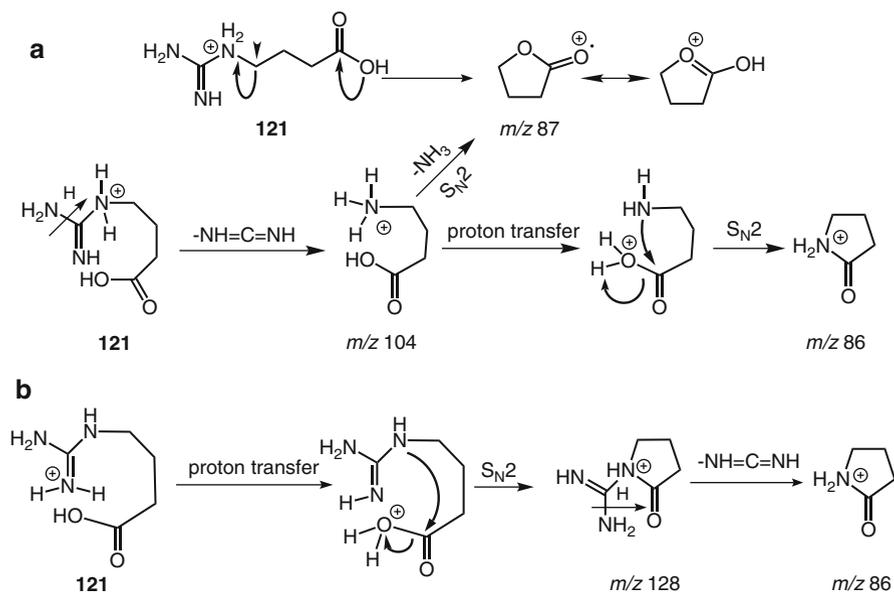
Scheme 2 Proposed fragmentation pathways for γ-guanidinobutyric acid (**125**) (*35*)

(−42 amu, NH=C=NH) from the ion of m/z 128 [MH−H$_2$O]$^+$ and a loss of water from protonated GABA (see Fig. 12a).

In contrast to acyclic guanidines, the loss of carbon diimine is a characteristic fragmentation reaction. The two compounds, cimipronidine (**60**) and cyclocimipronidine (**62**), were identified and their structures confirmed with extensive NMR data. In addition, a methyl ester of cimipronidine (**63**) as well as *N*-methylcyclocimipronidine (**108**) were identified by analogy with the fragmentation patterns of **60** and **62**.

2.3.2.3   Hydroxycinnamic Amides

Caffeic, ferulic, and isoferulic acids have been reported as major hydroxycinnamic acids in BC. The identification of amides and esters of ferulic/isoferulic acid (**72–75**, **100–102**, **105–106**, **128–131**) was enabled by their characteristic fragmentation pattern, which is dominated by the product ions of m/z 177, 149, 145, 117, and 89 originating from the ferulic/isoferulic acid moieties of the amide. Caffeic acid amides such as **71** produce a similar ion series at m/z 163, 145, 135, 117, and 89 (see Fig. 13c). The presence of a low abundance fragment ion of m/z 163 with the elemental composition of C$_9$H$_7$O$_3$ can be used to distinguish isoferulic acid from ferulic acid (*50*).
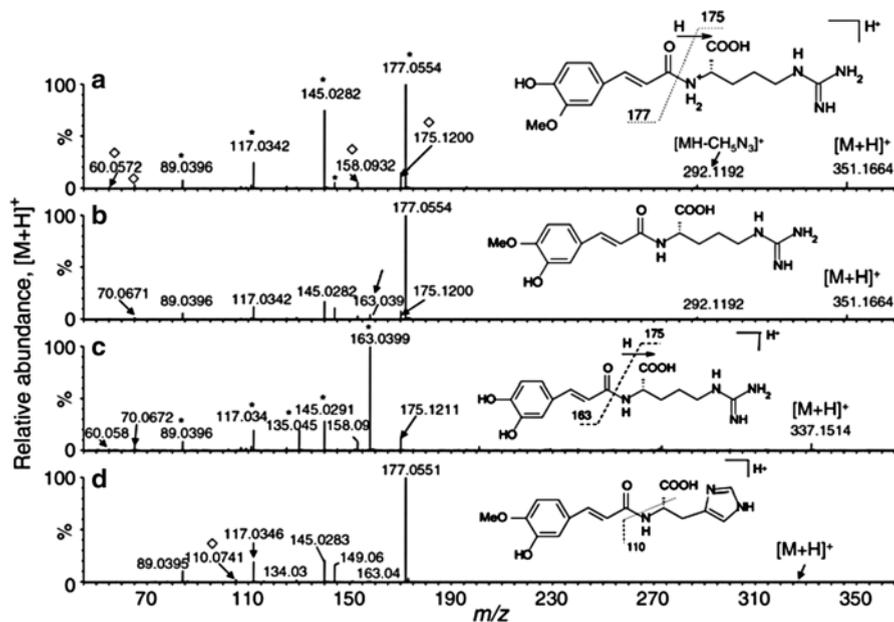
Fig. 13 Product ion tandem mass spectra of amides of hydroxycinnamic acids with amino acids: (**a**) feruloyl arginine (**72**), (**b**) isoferuloyl arginine (**73**), (**c**) caffeoyl arginine (**71**), and (**d**) isoferuloyl histidine (**78**). Ion series corresponding to the acid portion are labeled "*" for ferulic and caffeic acid in (**a**) and (**c**), respectively, while those corresponding to the amine portion are labeled "◇" for arginine and histidine in (**a**) and (**d**), respectively. From *Nikolić et al.* (*35*), with permission of Elsevier B.V.

Once the diagnostic ions from ferulic/isoferulic acid are observed in the product ion spectrum of an unknown compound, the amine portion of the amide can be deduced based on a database search of the elemental composition of the remainder of the molecule. As an example of this identification strategy, Fig. 13 shows the product ion spectra of compounds **72** and **73**, both with an elemental composition of $C_{16}H_{22}N_4O_3$. Both spectra show a typical ferulic/isoferulic acid amide fragmentation pattern, with **72** showing an additional peak at *m/z* 163. This suggested that **72** is an amide of ferulic acid, while **73** is an amide of isoferulic acid. The database search for the composition of the remainder of the molecule ($C_6H_{14}N_4O_2$) suggested that the amine portion is the amino acid arginine. Protonated arginine was observed at *m/z* 175, along with other ions originating from fragmentation of arginine such as ions of *m/z* 158, *m/z* 70 and *m/z* 60 (see Fig. 13a). The presence of the latter two ions indicated a free guanidino group and confirmed that the carboxylic acid was attached to the α-amino group rather than a guanidino nitrogen.

2.3.2.4 Choline and Betaine Alkaloids

Both choline and betaine alkaloids contain the quaternary nitrogen group, trimethylammonium. The characteristic fragment is a loss trimethylamine (-59 Da; Me_3N), which can be used to detect, specifically these two classes of alkaloids. For example, **104** displayed a characteristic loss of trimethylamine from the precursor ion of *m/z* 208 to form an ion at *m/z* 149, which can further fragment to lose CO_2 and produce an ion of *m/z* 105.

2.3.2.5 *Pictet-Spengler* Adducts with Tryptamine Derivatives

Compounds **96** and **99** eluted at 3.3 and 3.8 min, respectively, during the LC-MS and had identical elemental composition ($C_{12}H_{14}N_2O$) but very different fragmentation patterns (see Fig. 14a, b). The elemental compositions of several key fragment ions such as *m/z* 160, 159, 132, and 117 were the same as those observed in the product ion tandem mass spectra of N_ω-methylserotonin and serotonin (*33*), suggesting that **96** and **98** may have the same biosynthetic pathway through a *Pictet-Spengler* reaction. Accordingly, condensation of serotonin and N_ω-methylserotonin with formaldehyde was carried out under acidic conditions. By comparing the fragmentation patterns of these adduct ions with those of **96** and **99**, identical fragment patterns demonstrated that **96** and **99** indeed resulted from *Pictet-Spengler* addition of N_ω-methylserotonin with formaldehyde. Compound **96** was identified as 6-hydroxy-2-methyl-1,2,3,4-tetrahydro-β-carboline, while the structure of **99** was elucidated as 3,4,5,6-tetrahydro-7-hydroxy-5-methyl-1*H*-azepino[5,4,3-*cd*]indole by a combination of its NMR and MS data.

Fragmentation of **96** is dominated by retro *Diels-Alder* fragmentation to form an ion at *m/z* 160 (see Fig. 14a). As retro-*Diels-Alder* fragmentation is not possible for
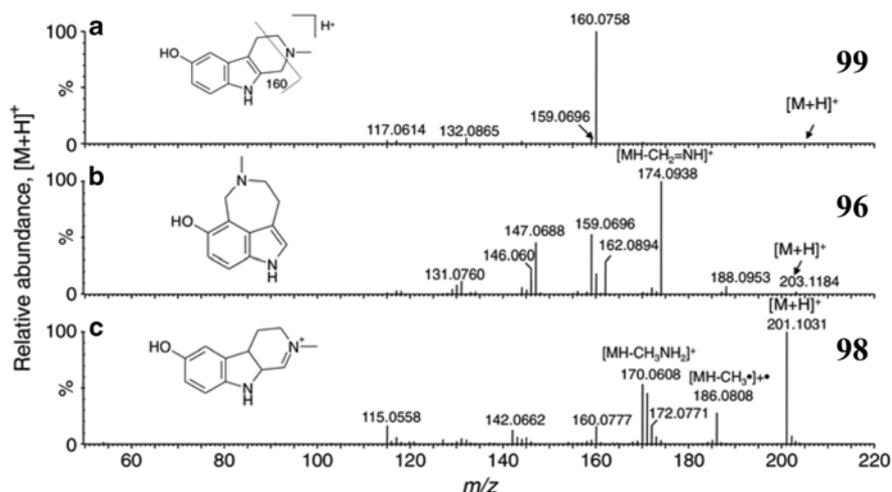
Fig. 14 Product ion tandem mass spectra of the *Pictet-Spengler* adducts of N_ω-methylserotonin and formaldehyde. From *Nikolić et al.* (*35*),with permission of Elsevier B.V.

99, fragmentation occurs by opening of the azepine ring, followed by elimination of methylene imine to form a base peak of *m/z* 174.0938 (see Fig. 14b). Both **96** and **99** originate from the same precursor species that cyclizes into either a six- or seven-membered ring (see Scheme 3), as has been demonstrated in studies of the reactions of serotonin and N_ω-methylserotonin with various aldehydes (*50, 51*).

The elemental composition of **98** ($C_{12}H_{12}N_2O$) is two hydrogens less than that of **96** and **99** ($C_{12}H_{14}N_2O$), suggesting a dihydro-β-carboline structure. The loss of a methyl radical (*m/z* 186), along with the fragment ion of *m/z* 170, [MH-CH$_3$NH$_2$]$^+$, indicated that the N(2) nitrogen on the β-carboline ring is methylated. Based on biosynthetic considerations, the most likely position of the double bond should be between 1 and 2. This structure was confirmed by comparison of retention time and fragmentation pattern with an authentic sample of *N*(2)-methyl-6-hydroxy-3,4-dihydro-β-carboline.
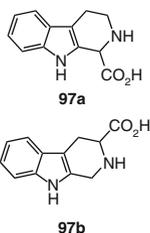
The product ion spectrum of compound **97** was dominated by an ion of *m/z* 144 with the elemental composition ($C_{10}H_{13}N_2$), corresponding to protonated
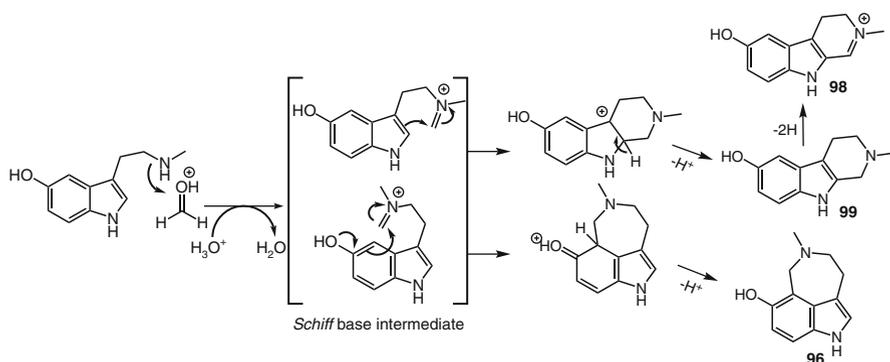
Scheme 3 Proposed mechanism of formation of **96**, **98**, and **99** (*143*)

tryptamine. The fragmentation of the product ion *m/z* 144 showed an identical pattern to authentic tryptamine, suggesting that this compound is a tryptamine derivative. The neutral loss of iminoacetic acid ($C_2H_3NO_2$) combined with database searching suggested that **97** might be a tetrahydro-β-carboline carboxylic acid. As both 1- and 3-substituted isomers are known, both analogs **97a** and **97b** were synthesized and compared with **97** in order to confirm the correct regioisomer as **97b**.

2.3.2.6 Other Alkaloids

The fragment patterns of other alkaloids, including benzylisoquinoline (*52*), aporphine (*53*, *54*), protoberberine (*55*, *56*), and protopine alkaloids (*47*) as well as other miscellaneous classes (*57–60*) have been discussed in detail elsewhere. Verifying their presence in *A. racemosa* (*35*) was mostly based on spectral database searching and comparison with appropriate authentic standards.

3 Fingerprinting

Recently, the concept of "fingerprinting" has been used increasingly in herbal medicines for their chemical/biological profiling, botanical identification, and quality standardization. The basic theory is that the individual plant samples have unique genetic and metabolomic profiles, which can be converted into "fingerprints" and used for identification. Commonly used techniques for depicting these characteristic profiles include DNA sequencing, chromatographic, and spectroscopic analysis, and a combination of these. As the resulting profiles are usually comprised of large and complex datasets, chemometric approaches such as classification, pattern recognition, and clustering can be employed to enable or facilitate data interpretation and comparison.

Fingerprinting of *A. racemosa* has been reported using a variety of chemical and biological techniques. These have focused mainly on two classes of compounds, the triterpenes and the phenolic constituents, which are frequently considered as the major bioactive substances in *A. racemosa*. *Verbitski et al.* developed an efficient and economical approach using thin-layer chromatography (TLC) combined with bioluminescence, which provided characteristic patterns corresponding to toxicity profiles for *A. racemosa* as well as many other *Actaea* species (*61*). This technique may not only be used to examine *A. racemosa* adulterants, such as blue cohosh (*Caulophyllum thalictroides*), but also to detect unknown contaminants. However, TLC is still less frequently used for chemical fingerprinting due to its limited dynamic range and/or resolution, relatively low sensitivity, and limited specificity. In comparison, HPLC-based methods are more popular, but not necessarily always more specific or of higher resolution. UV, ELSD, and MS detection are commonly used in online hyphenated HPLC separation methods. Of these three detection methods, UV is very appropriate for the detection of phenolic constituents in *A. racemosa* (*62*, *63*). However, due to the lack of UV pharmacophores in most of the triterpenes, they are better detected by ELSD (*64–66*) or MS. Compared to UV and ELSD, MS is much more powerful for the identification and quantification of both triterpene and phenolic constituents, due to higher sensitivity and selectivity, but also by providing more structural information. A number of LC-MS techniques have been employed for the fingerprint profiling of *A. racemosa*. Some of the examples include LC/APCI-MS (*19*, *63*, *65*), LC/TIS-MS (*67*), and LC/EI-MS/MS (*68*). Instead of using a single detection method, *He et al.* reported a comprehensive approach based on HPLC-PDA/APCI-MS/ELSD for the fingerprinting of 10 *Actaea* species, specifically, *A. acerina*, *A. americana*, *A. biternata*, *A. dahurica*, *A. foetida*, *A. heracleifolia*, *A. japonica*, *A. racemosa*, *A. rubifolia*, and *A. simplex* (*62*). The combined use of three spectroscopic methods provides increased reliability and versatility for the detection, identification, and quantification of both triterpenes and phenolic constituents in these plants. As a result, the HPLC profiles from multiple detectors provide more detailed fingerprints of the plant samples, and thus improves the overall accuracy of plant identification by chemical fingerprinting.

In addition to chemical methods, a few biological methods have been described for the fingerprinting of *A. racemosa*. *Zerega et al.* first reported the use of a DNA fingerprinting technique, Amplified Fragment Length Polymorphism (AFLP), for the differentiation of *A. racemosa* from three other closely related species, *A. pachypoda*, *A. cordifolia*, and *A. podocarpa* (*69*). In using four AFLP primer combinations, from these four species altogether 262 unambiguous DNA fragments were generated, of which one was unique to *A. racemosa*. This characteristic DNA marker was used as a fingerprint to authenticate two commercial *A. racemosa* products. In a later report, *Motley et al.* (*70*) optimized the AFPL analysis and demonstrated that the use of only two-primer combinations was sufficient to identify each species. Principal component analysis (PCA) of the AFPL results enabled the identification of the geographical region from which an *A. racemosa* specimen was collected. Furthermore, *Baker et al.* adapted a DNA barcoding methodology to identify

unambiguously *A. racemosa* in dietary supplements (*71*). According to the Consortium for the Barcode of Life (CBOL), two protein coding regions, *matK* and *rbcL*, can be used as core plant DNA barcodes, and two non-coding regions, *nrITS* and *psbA-trnH*, can be utilized as supplemental markers. Four regions from the DNA samples extracted from *Actaea* plants were sequenced after PCR amplification, and two *matK* nucleotide sequences were found to be specific for *A. racemosa*. Using these two markers, *A. racemosa* was correctly distinguished from other *Actaea* species. Further application of these markers was carried out to examine the adulteration of *A. racemosa* dietary supplements by other *Actaea* species. It should be noted that in mainland China, several other species are sold as "Black Cohosh". Some of these species do not belong to the genus *Actaea*, but include *Vernonia aspera*, which is known locally as *hei-sheng-ma* and is sometimes translated as "Black Cohosh" (*4*).

4 Names and Origin

As alluded to in the Introduction, "Black Cohosh" has been used with several meanings implied. Today, perhaps the most generally intended meaning of "Black Cohosh" is that of an extract of the underground parts (roots and rhizomes) of *A. racemosa*—but this is often only implied and not specified (note the definitions for this work, made in the Introduction). However, the term is also used to mean the dried roots and rhizomes, themselves, and occasionally even for the whole plant. The plant was first described by Europeans as *Christopheriana facie* in 1690 in *Leonard Plukenet*'s *Phytographia*, and later was given the name *Actaea racemosa* by *Linnaeus* in 1751. Despite this unassailable authority, the name was changed to *Macrotrys racemosa,* which morphed into *Macrotys racemosa* and was subsequently changed to *Cimicifuga racemosa*. The last name was widely used until recently when on the basis of 26SrDNA comparison, botanists reorganized the genera, *Actaea, Cimicifuga,* and *Souliea,* and the last two genera disappeared with all of their species being incorporated into *Actaea* (*72*). There are now 28 different species of *Actaea* spread across the entire Northern Hemisphere, with eight endemic to North America and the other 20 occurring mainly in Asia, with three also found in Europe. Despite the predominant use of "Black Cohosh" in the form of root/rhizome extract, even this assignment is not definitive as various types of extracts are sold; 60–80% ethanolic as well as 40% *iso*-propanolic and other extracts have been used in the search for metabolites or in pharmacological studies. This makes comparison of results from studies, and particularly meta-analysis of clinical trials very difficult and sometimes even impossible. Notably, one of the most widely used commercial preparations, Remifemin®, has been reformulated from an ethanolic extract to a 40% *iso*-propanolic extract.

Most plant harvesting (>90%) is from native wild growth (*4*), and this leads to some concern for sustainability, and to both accidental and deliberate adulteration. *A. podocarpa* plants resemble those of *A. racemosa*, and the two species share a

fairly common habitat. The traditional Chinese medicine, *Sheng-Ma*, described as *Rhizoma Cimicifugae*, includes an extract of the roots and rhizomes of *A. heracleifolia, A. dahurica, A. foetida* and, less frequently, *A. simplex* and *A. yunnanensis.*

One example of highly probable adulteration is the recent discovery of two new triterpenes (**135** and **136**), which were named "cimipodocarpaside" and "isocimipodocarpaside", respectively, from a purchased, unauthenticated[1] sample of "Black Cohosh". Prior to this report, all triterpenes isolated from *Actaea* having the 9,10 bond cleaved had been described from *A. podocarpa.* It seems likely that this purported sample of Black Cohosh contained extracts of species of *Actaea* other than *racemosa* (*73*, *74*).
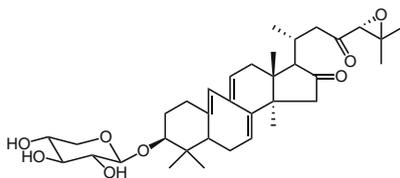
135 (7,8,9,11,10,19-hexadehydro-(24*R*)-24,25-epoxy-16,23-dioxo-9,10-*seco*-3-*O*-β-D-xylopyranosyl-actanoside)
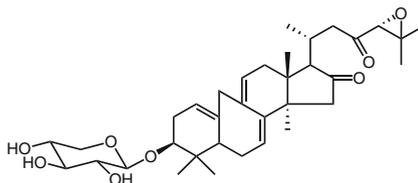
136 (1,10,7,8,9,11-hexadehydro-(24*R*)-24,25-epoxy-16,23-dioxo-9,10-*seco*-3-*O*-β-D-xylopyranosyl-actanoside)

Many commercial preparations containing BC are mixtures with extracts of other medicinal herbs, such as red clover, soy bean, magnolia bark, ginkgo, hops, damiana leaf, *St. John*'s wort, *Scutellaria lateriflora, Valeriana officinalis, Passiflora incarnata,* chaste tree berry, China root, Chinese bupleurum root, codonopsis root, ginger root, lovage root, rehmannia root, white peony root, horse chestnut seed, and licorice root. However, a number of products purport to be an extract of *Actaea racemosa* roots/rhizomes, unadulterated with other herbs. These include Remifemin®, Natrol®, Spring Valley®, and CR BNO 1055.

Jiang et al. (*75*) analyzed the air-dried roots/rhizomes of 15 different species of *Actaea*, eight endemic to North America and seven to Asia, sampling from 1 to 15 samples per species. Their analysis established some clear marker compounds: cimiracemoside F (**38**) (*76*, *77*) was present in all samples (10) of *A. racemosa*, and only in one other sample allegedly from *A. pachypoda.*

---

[1]An attempt to ascertain authentication resulted in a letter from the Chinese supplier stating that this was Black Cohosh (personal communication from one of the authors).
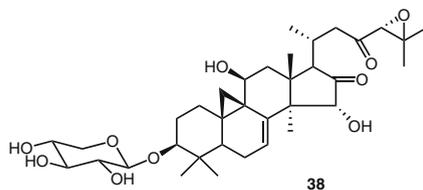
38

Cimifugin (**137**) (*78*) was present in all 27 samples from an Asian species, but only in trace quantities in the four samples from three North American species, two of which are endemic to the Pacific Northwest. Hence, the lack of **38** and the presence of **137** in a commercial sample is compelling evidence that the purported "Black Cohosh" is, in fact, from an Asian species and is not authentic material. The same group (*75*) had previously carried out an investigation of 11 commercial products available over-the-counter in the New England area and concluded that four were from Asian species, although one of these may have contained some *A. racemosa*.
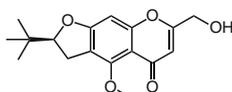
137 (cimifugin)

Perhaps one of the more unfortunate findings was that (12*R*)-12-acetoxy-(24*R*,25S)-24,25-epoxy-(26*R*/S)-26-hydroxy-3-*O*-*β*-D-xylopyranosylacta-(16S,23*R*)-16,23;23,26-binoxoside (actein) (**24**) and (12*R*)-12-acetoxy-(24*R*,25*R*)-24,25-epoxy-3-*O*-*β*-D-xylopyranosylacta-(16S,23*R*)-16,23;23,26-binoxoside (23-*epi*-26-deoxyactein) (**29**) were found in good quantity in all species examined, with the exception of *A. cordifolia*, *A. podocarpa*, and *A. laciniata.* These two compounds have been, and continue to be, common markers used to calibrate the potency and supposed authenticity of commercial BCEs. Commercial preparations occur as capsules, tablets or solutions, most commonly derived from BCEs, but some are commonly found to contain ground dried BC. Reported extraction solvents vary from 40% isopropanol to 58–79% ethanol and 80% methanol, and all these products/materials have very different triterpene contents (*75*) and different potencies with respect to binding to the 5-HT$_7$ subtype of the serotonin receptor (*79*). The latter can be explained readily with differences in solubilities of the serotonergic *in vitro*-active principle, N_ω-methylserotonin (**61**) in the different extraction solvents.
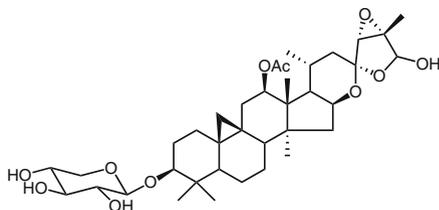
24 ((12*R*)-12-acetoxy-(24*R*,25S)-24,25-epoxy-(26*R*&S)-26-hydroxy-3-*O*-*β*-D-xylopyranosylacta-(16S,23*R*)-16,23;23,26-binoxoside)
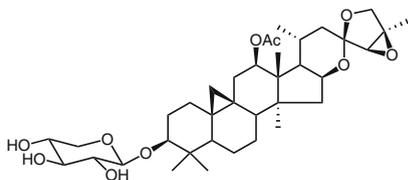
29 ((12*R*)-12-acetoxy-(24*R*,25*R*)-24,25-epoxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside)

5   Pharmacology

5.1   *Estrogenic Activity*

A major impediment to the rigorous comparison of results from different studies evaluating potential estrogenic effects of BCEs is the fact that, although it is usually possible to ascertain that "Black Cohosh" is being used to describe the roots and rhizomes of *A. racemosa* or an extract thereof, the nature of the extract varies widely through the use of different alcohols, methanol, ethanol, and isopropanol, with or without up to 80% water, and made from room temperature to under reflux. In some cases, a commercial preparation such as Remifemin® has been used, but as mentioned earlier even this product has changed its formulation from an ethanolic extract to an *iso*-propanolic one.

It may seem very logical, as BCEs have been used to alleviate the adverse symptoms of menopause and menstruation, and as this extract is rich in triterpenes with a clear structural relationship to steroidal hormones, that one or more constituents of BCE probably has phytoestrogenic activity. However, attempts to define such a mode of action have been inconclusive. Estrogenic activity was assessed in a number of different plant extracts, including BCEs using several assays by *Liu et al.* (*80*). These assays included binding to the α- and β-estrogen receptors, estrogenic activity as evidenced by induction of alkaline phosphatase, and/or up-regulation of progesterone receptor mRNA in *Ishikawa* cells (endometrial), as well as up-regulation of presenelin-2 in S30 breast cancer cells. Whereas hops, red clover and chasteberry extracts showed significant activity in the estrogen binding and the Ishikawa cell assays, and although extracts of *Angelica sinensis,* licorice, Asian and American ginseng showed modest activity in the S30 cell assay, a methanolic BCE showed no activity in any of these assays at 20 μg/cm^3. Subsequently, *Einbond et al.* (*81*) reported that a BCE inhibited the growth of MCF7 human breast cancer cells. However, the most potent compound was 25-acetoxy-7,8-didehydro-(12*R*)-12-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside (**138**), obtained from *C. acerina*. This compound inhibited the growth of MCF7 (ER + Her2) human breast cancer cells with an IC_{50} of 3.2 μg/cm^3, whereas actein showed an IC_{50} of 5.7 μg/cm^3. For both compounds, this inhibition was increased by further transfection of the cells with Her2.

138
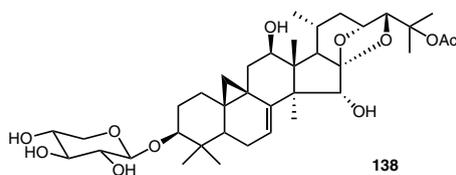
In addition to the triterpenes, the isoflavonoid, formononetin, a known phytoestrogen, has been reported as a constituent of BC. Initially reported in 1985 (*82*), its presence was disputed by two groups (*64*, *83*), both in 2002. The 2002 results were refuted in 2004 (*84*) with the claim that a more accurate analysis could detect formononetin in dried BC at levels of 3.1–3.5 µg/g. In 2006, the *Kennelly* group, after extending the limits of detection of formononetin to 60 ng/g dry weight, was still unable to detect the compound in any of 13 samples of BC (*65*). In summary, presently there is no chemical evidence, and also no chemotaxonomic indication, for direct *α*- or *β*-estrogenic activity caused by any BCE or components thereof.

5.2 Prevention of Bone Loss

Seidlová-Wuttke et al. (*85*) have made the case that BCE (BNO 1054, a cold 50% ethanol/water extract) is actually a SERM (selective estrogen receptor modulator), having little or no effect on the uterus in comparison with estradiol-17*β* (E-17), but inhibiting LH secretion in acute treatment of ovariectomized rats. On chronic treatment (3 months as a food additive), both E-17 and BCE counteracted the metaphyseal bone loss of the femur seen in control untreated ovariectomized animals, although BCE at 400 mg·kg^{-1}·day^{-1} was less effective than E-17 at 0.5 mg·kg^{-1}·day^{-1}. Some correlation of these effects was seen in gene expression for the Type 1 procollagen α-chain, osteoprotogenin, osteocalcin, and tartrate-resistant-α-phosphatase Type 5 genes. A fat deposit seen in the CT scans of the tibias of chronically treated rats was quantified and significantly increased in untreated animals. This increase was almost equally reduced (>50%) in rats treated with E-17 at 0.5 mg·kg^{-1}·day^{-1} and BCE at 100 and 400 mg·kg^{-1}·day^{-1}. It was shown that Remifemin® at 4.5 mg of triterpene content·kg·day^{-1} and raloxifene at 3 mg·kg^{-1}·day^{-1} when administered orally to ovariectomized rats were both equally effective at lowering urinary levels of the known markers of bone loss, pyridoline, and deoxypyridoline. Moreover, the authors were able to maintain a high level of resistance to bone fracture in these rats.

A mechanistic basis for these beneficial effects on bone structure was provided by *Qiu et al.* (*86*) who showed that 25-acetoxy-(15*R*)-15-hydroxy-3-*O*-*β*-D-xylopyranosylacta-(16*S*,23*R*,24*S*)-16,23;16,24-binoxoside (25-*O*-acetylcimigenol xylopyranoside) (**5**) inhibited the RANKL or TNF-*α* pathways to osteoclastogenesis with an IC_{50} of approximately 5 µ*M*, while other triterpenes in BCE [cimigenol (**16**), actein (**24**), and cimiaceroside B (**40**)] were inactive or less potent with IC_{50} values of 25, 42, and 45 µ*M*, respectively.

Ruhlen et al. (*87*) found that 23-*epi*-26-deoxyactein (**29**) decreased cytokine-induced NO production in murine microglial cells. However, whole CimiPure® (a BCE containing 2.5% triterpenes) increased this production.

Kim et al. (*88*) showed that a cold methanol BCE had a strong anti-allergenic effect in several assays, including a murine skin sensitization assay (the local lymph node assay) in which the extract was not an allergen itself. When the extract was administered to rats twice daily for three days, it showed a strong dose-dependent inhibition of anti-IgE-induced passive cutaneous anaphylaxis. *In vitro*, the extract inhibited the histamine release induced by compound 48/80 in rat peritoneal mast cells. In human leukemia mast cells the mRNA for the inflammatory cytokines, IL-4, IL-5 and TNF-α were less induced by PMA and A23187 in the presence of the extract, but, however, this effect had an inconsistent dose-dependency.

Lee et al. (*89*), when working with extracts of *A. heracleifolia*, showed that 25-*O*-acetylcimigenol-3-*O*-β-D-xylopyranoside (**5**) was a strong inhibitor (*IC*$_{50}$ 7.7 μ*M*) of the complement pathway. This compound is present in BC at about two-thirds the concentration of 23-*epi*-26-deoxyactein (**29**) (*81*).

5.3 Potential Anticancer Activity

Burdette et al. (*90*) have shown that a BCE showed efficient free radical (DPPH) scavenging effects, and with bioassay-directed fractionation ascribed this activity to nine phenolic compounds, methyl caffeate, ferulic acid, isoferulic acid, fukinolic acid, cimicifugic acid A, cimicifugic acid B, cimicifugic acid F, cimiracemate A, and cimiracemate B. These authors went on to show that five of these compounds plus caffeic acid were able to reduce menadione-induced DNA damage in S30 human breast cancer cells, damage that is known to be initiated by reactive oxygen species. Actein (**24**) inhibited the growth of p53 positive HepG2 liver cells *in vitro* (MTT assay) with an *IC*$_{50}$ of 27 μg/cm^3 (*81*).

Mimaki et al. (*7*) isolated from a hot methanolic BCE, (24*S*)-24-acetoxy-3-*O*-α-L-arabinopyranosyl-(15*R*)-15-hydroxyacta-(16*S*,23*R*)-16,23;16,25-binoxoside (**44**) and its xylopyranosyl analog. These compounds had no effect on the ATCH secreted from AtT20 cells (mouse anterior pituitary tumor cells) by themselves, but each significantly increased the effect of corticotrophin-releasing factor (CRF) on the ATCH release by these cells.

5.4 Stress Relief

Using single doses of 7.14 and 35.7 mg·kg^{-1}, *Einbond et al.* (*91*) have studied the effect of orally administered actein (**24**) on stress and statin-associated responses in *Sprague–Dawley* rats. The study also included measurements of the blood levels, which peaked at 2.4 μg/cm^3 at 6 h for the high dose and fell to 0.1 μg/cm^3 at 24 h.

The authors saw statistical differences in the transcription of 297 and 1325 genes at 6 and 24 h, respectively, compared with control animals (gavage with water). The genetic effects included the down-regulation of erythropoietin, CYP2C, and ATP synthase genes at 6 h, and suggested that the primary effects of actein (**24**) "may be on hypoxia and the stress response and mitochondrial phosphorylation." There was a weak signature match to genes involved in the effects of statins on cholesterol biosynthesis, and sections of the fatty acid biosynthesis pathway were significantly down-regulated. The study confirmed the effects on cholesterol biosynthetic pathways by showing that actein (**24**) inhibited the proliferation of p53 positive HepG2 liver cells. Interestingly, the reported IC_{50} values were identical with those determined by *Burdette et al.* 7 years earlier (*90*).

Nadaoka et al. (*92*) extended their earlier studies on the effect of BCE on stress in mice by a small, but well designed, placebo controlled, randomized, double blinded, cross-over study in healthy adult humans. In this, they monitored both physiological and psychological responses to stress induced by mental arithmetic tasks beyond capable completion in a limited time period. Salivary cortisol and chromogranin-A were measured before administering the drug, and during and immediately after the task, and then following a rest period. Both indicators rose significantly during the task and returned to near normal values by the end of the task. There was no significant difference between placebo and treated groups in the cortisol levels, but there was a major amelioration in the rise of chromogranin-A levels in the treated group compared with the controls. Psychological evaluation was done with a visual analog scale and use of the Japanese version of the State-Trait Anxiety Inventory (*93*).

In a separate, but similar, EEG monitored experiment, topographical changes in the alpha waveband from both the left and right occipital areas were strongly decreased during the task, and only the treated groups showed a tendency to recover after the rest period, although these failed to reach normally accepted statistical significance ($P = 0.06$ for left and 0.07 for right occipital areas). The authors proposed that BCE could "be useful for the prevention and treatment of stress related disorders."

More recently, mode of action studies on BCEs have concentrated on CNS effects and this emphasis owes its origin to the isolation and characterization of N_ω-methylserotonin from a cold methanolic BCE and its very potent 5-HT$_7$ receptor binding ($IC_{50} = 23$ p*M*), ability to induce cAMP ($EC_{50} = 22$ n*M*), and to block serotonin uptake ($IC_{50} = 490$ n*M*). These findings created a compelling argument that at least some of the effects of BCE may be *via* the CNS. The authors estimated that a 120 mg dose of BCE would contain 3.7 μg of N_ω-methylserotonin (*33*).

Following up on the CNS-active constituents (*11*), *Cicek* and associates carried out a bioassay-guided fractionation of an exhaustive *Soxhlet* methanolic BCE using an assay to identify synergistic modulators of GABA in the induction of chloride currents. The authors isolated 11 cycloartane glycosides, of which only actein (**24**), 23-*O*-acetylshengmanol-3-*O*-β-D-xylopyranoside (**32**), cimigenol-3-*O*-β-D-xylopyranoside (**1**), and 25-*O*-acetylcimigenol-3-*O*-α-L-arabinopyranoside (**6**) increased the GABA-induced chloride currents, with compound **32** doing so by

700%. This compound was able to increase chloride currents by 150% in the absence of GABA. Enzymatic hydrolysis of **1**, **24**, and **32** yielded the corresponding aglycones, which were only capable of weak enhancement of GABA-induced chloride currents.

5.5 Hepatotoxicity

In 2002, the United States Pharmacopeia, Inc. (USP) Dietary Supplements Information Expert Committee (DSIEC) recommended a GRAS classification for BCEs. However, by the end of that year, the first report of association of BCE with hepatotoxicity occurred when an article entitled "Black Cohosh and other Herbal Remedies Associated with Acute Hepatitis" described six instances occurring from 1996 through 2001 in patients taking some herbal complementary medicine, of which only two included a BCE. One of these two patients had used a BCE for only one week as a sole treatment for menopausal symptoms. Her hepatic failure was acute and severe, and she underwent a liver transplant a week after presentation. The other was also taking skullcap and valerian, both had been reported previously to be hepatotoxic. This report was immediately criticized by *Vitetta et al.* (*94*) on the grounds that there was no data on the source or authentication of the BCE, nor was viral hepatitis adequately excluded as the cause. Nonetheless, another report by *Lontos et al.* (*95*) appeared almost concurrently with *Vitetta*'s criticism and cited a single case of a woman who had taken a BCE for three months and developed severe chronic hepatitis, which worsened over the month following cessation and required a liver transplant. In this case, the source of the BCE was given, but no authentication was supplied. These and other reports worldwide led the Australian Therapeutic Goods Administration to review 47 worldwide instances of BCE-associated liver toxicity reported by 2005 and require a label containing a cautionary message ("Warning: Black cohosh may harm the liver in some individuals. Use under supervision of a healthcare professional"; http://www.tga.gov.au/safety/alerts-medicine-black-cohosh-070529.htm), on all BCE-containing commercial preparations.

Health Canada-Santé Canada (HC-SC) issued a statement on BCEs in which they advised users to consult a healthcare practitioner if any liver trouble developed (http://www.hc-sc.gc.ca/dhpmps/alt_formats/hpfb-dgpsa/pdf/prodnatur/mono_cohosh-grappes-eng.pdf). By April 2009, HC-SC had received six case reports of purported BCE/hepatotoxicity association. On examination, four of these were associated with preparations from a single supplier, who withdrew the product as it actually did not contain any *A. racemosa*. No analysis of the product was available in the other two cases.

The Committee on Herbal Medicinal Products (HMPC) of the European Medicines Agency, after studying 31 European cases, 11 world-wide case reports, and 15 clinical trials concluded that almost all cases reporting hepatotoxicity and BCE use were poorly documented, and that three cases could be classified with

causal relationship as "possible" and two as "probable". Most cases were judged as unlikely to have any relationship between BCE consumption and hepatotoxicity, or were simply so poorly documented that they had to be excluded from consideration. The 2007 HMCP report recommended that any patient taking BCE should contact their physician immediately on developing any signs of liver injury (http://www.emea.europa.eu.int/pdfs/human/hmpc/26925806en.pdf).

The same year, the USP DSIEC, after reviewing 30 reports of liver damage following use of products containing BCE, decided that preparations containing BCE should be classified as Class 2 hepatotoxic agents (*i.e.* a possible but not probable cause) and carry a label advising users to discontinue taking the preparation and consult a healthcare practitioner upon any signs of liver trouble (*96*).

In the same year, *Borrelli et al.* (*97*) published a "systematic review of adverse events" associated with BCEs, updating an earlier review. They surveyed clinical trials involving 1,522 patients, post-marketing surveillance of 2,691 patients, and case reports covering 18 patients for adverse effects. Women from 24 to 84 years of age were involved, and they were treated for periods ranging from 1 week to 1 year. There was no hepatotoxicity reported in either the clinical trials or the post-marketing surveillance patients. Among the case reports, seven involved liver problems, but, on examination, only one of these could be assigned probable cause from a BCE and two others, a possible cause. The relative safety of BCEs, specifically with respect to breast cancer patients, has been emphasized in a 2010 review (*98*).

In 2011, *Teschke et al.* (*99*) reviewed the literature on case reports of BCE-associated hepatotoxicity, with careful reference to the definition of herb-induced liver injury as defined by the CIOMS (Council for International Organizations of Medical Sciences), and also for the identification, quality, and authentication of the particular herbal preparation used. Their conclusion was that there is currently no evidence for a causal relationship between BCE and any of the case reports of purported hepatotoxicity.

Virtually simultaneous with *Teschke's* review, *Naser et al.* (*100*) published the results of a meta-analysis of randomized controlled clinical trials of Remifemin® or Remifemin® plus (*i.e.* a 40% *iso*-propanol extract of *A. racemosa* roots/rhizomes without or with *St. John*'s wort extract, respectively) in healthy pre- and post-menopausal women treated with 40 mg (conventional dose) to 128 mg (high dose) of extract daily for three or six months. Five different trials comprising 1,108 women (557 treated and 551 on placebo) met the inclusion criteria. Eighty-eight women dropped out of the trials, none as a result of liver problems, and all except one of whom had enzyme level data collected as they left the study. Liver function was assessed by measuring aspartate aminotransferase, alanine aminotransferase, and β-glutamyltranspeptidase on entering and at the completion of the trial. There was no significant difference between treated and placebo values or other suggestion of liver toxicity as a result of BCE treatment.

The generally accepted conclusion is that there is no convincing evidence that BCEs cause hepatotoxicity; however, the possibility exists that it may do so in rare idiosyncratic cases. Hence the U.S., Australia, Canada, and the U.K. all require cautionary warnings on all preparations containing a BCE to the effect that patients

using these preparations should consult a healthcare provider in the event of any symptoms suggestive of liver malfunction.

Animal studies, predominantly in rats, investigating the hepatotoxic potential of BCE have given mixed results. *Lüde et al.* (*101*) observed microvesicular steatosis in rats following doses of 1000 mg·kg^{-1} (several orders of magnitude above the normal clinical dose). *In vitro* cytotoxicity with HepG2 cells was observed at 75 µg/cm^3, and mitochondrial β-oxidation was impaired at 10 µg/cm^3. The authors concluded that their studies were "compatible with idiosyncratic hepatotoxicity as observed in patients". *Mazzanti et al.* (*102*) found a reduction in liver GSH levels in male rats given 300 mg·kg^{-1}·day^{-1} for 30 days. However, they concluded that this dose was "quite safe in rats".

Campos et al. (*103*) reported a multidimensional study with female *Wistar* rats comparing groups aimed at mimicking conditions in post-menopausal women, *i.e.* ovariectomized and ovariectomized with hypertension induced *via* the 2 kidney, 1 clip model. A subsection of the latter group was treated daily with 0.6 mg/kg with an unspecified BCE. The treatment increased the reactive oxygen species (ROS) generated by mitochondria as measured by DCF generated from DCFH-DA. The GSH levels were significantly decreased in all experimental groups compared with levels in normal rats. The glucose-6-phosphate dehydrogenase activity was significantly less in the two untreated ovariectomized groups than in normal rats, but this effect was significantly moderated by treatment with BCE. Lipid peroxidase levels, as assessed by levels of thiobarbituric acid reactive substances, was significantly increased only in the BCE treated group. The authors concluded from this study that BCE may make women more susceptible to toxic effects of other drugs as GSH levels may be depressed to a greater extent by high doses or prolonged treatment. However, they admitted that these effects were not seen in their study.

6   Clinical Trials

As this contribution demonstrates, BCEs have been the subject of extensive research due in part to a long history of use, which can be traced to centuries-old Native American healing traditions. However, even as new technologies and analytical techniques have become available, teasing out the active principles and mechanism of action as applied to BCE's clinical benefits remains a challenge to investigators. To date, there have been more than forty clinical studies of BCE, used to address menopausal symptoms, reported in the medical literature. A comprehensive review and summary of BCE studies dating from 1957 to 2009 by *Fabricant et al.* has been published in the *Encyclopedia of Dietary Supplements* (*104*). The early studies dating from the 1950s appear in the medical literature of Germany, where botanical medicines were well established as part of pharmaceutical armamentarium available to physicians. The majority of these early interventions were open studies in the style of serial case reports, most having a study length of about 12 weeks. Physicians

demonstrated their patients' responses to the BCE intervention, generally reporting improvements in menopausal symptoms with few and mild adverse events.

Reviewers of the earlier studies—those published in early 2000 and before—generally concluded that BCE was a safe intervention for short-term treatment of menopausal symptoms (*105–108*). During the early to mid-2000s, case reports began to appear in the medical literature impugning the safety of BCE, suggesting a link between its use and liver toxicity (see also Sect. 5) (*95*, *109–111*). Reviewers of these adverse responses note that both the evidence for causality and the case review methodology have been lacking (*112*, *113*). The peers emphasize the importance of thorough clinical evaluations in association with adverse event reporting that adhere to inclusive evaluation guidelines. Nevertheless, the placement of warning labels on BCE botanical dietary supplements has been both proposed and challenged (*114*, *115*). Interestingly, results of two separate year-long randomized, double-blinded, placebo-controlled clinical trials, each of which included at least one BCE arm, revealed no liver toxicity during the course of these trials, which were conducted in the mid-2000s. One is the University of Illinois at Chicago (UIC)/National Institutes of Health (NIH) Botanical Center trial, in which post-menopausal women took 246 mg of a BCE (75% ethanol extract) daily for one year (*116*). The investigators reported that there was no difference between BCE and placebo for any of the monitored safety parameters; significantly, no evidence of hepatotoxicity was uncovered. Similarly, participants in the BCE arm of the Group Health trial took 160 mg of a BCE (a 70% ethanol extract) daily for the one-year duration of the trial (*117*). In a second arm of the same trial, participants took 200 mg daily of a BCE (a 4:1 water and alcohol extract), which was combined in a multi-botanical mixture. The authors note that there were no statistically significant differences between the four arms of the trial (two of the four arms contained BCE) and placebo. Any severe adverse responses were too few to make meaningful comparisons between groups, but importantly hepatotoxicity was not included among them.

Since the late 1980s, clinical interventions of BCE have evolved to randomized, placebo-controlled clinical trials with mixed results, which can be found in the *Fabricant et al.* summary (*104*). Some trials have shown benefit in decreasing menopausal symptoms, while others have shown that BCE is not statistically different from placebo. Different theories have been proposed for these dichotomous results. Some of the key points of discussion and related insights are addressed in the following.

The first aspect relates to clinical methodology. In examining the clinical trial design of these and other menopausal intervention studies, it should be noted that assessing the benefit of drug or botanical interventions has included the use of questionnaires. Often referred to as climacteric scales, these subjective instruments are used to rate menopausal symptom severity using a numerical score. Examples used in BCE clinical trials are the *Kupperman* Index, Menopausal Rating Scale (MRS), and *Greene* Climacteric Scale. In the same manner that diaries of hot flashes are often employed in clinical trials, where patients are asked to report the frequency and severity, the climacteric scales rely on self-assessment and self-reporting. Additionally, clinical investigators have developed and increasingly emphasized

objective measures to identify the impact of BCEs on physiology, including measuring hormone levels (*e.g.* serum estradiol, follicle-stimulating hormone, luteinizing hormone, steroid hormone-binding globulin), vaginal cytology, *etc.* Also of interest is a technological innovation for measuring hot flashes objectively, the hot flash monitor, which has been introduced in recent clinical trials (*118*).

The second broader aspect, which might contribute to the dichotomous clinical outcome with BCE preparations, relates to the underlying biological targets. Coincident with the utilization of broader physiological assessment tools used in BCE studies, a consensus has developed among many investigators that BCE does not act as a classic phytoestrogen as presumed earlier, but rather may mimic selective estrogen receptor modulator effects on the bones and/or act on the CNS (*79*, *119–121*). Subsequent *in vitro* and human studies have signaled the activity of BCE on serotonin, *mu*, as well as GABA$_A$ receptors, with the suggestion that one or all of these biological interactions may function to bring about the diminution of menopausal symptoms, as reflected in many BCE clinical studies (*11*, *33*, *122*, *123*).

While assessing the biological impact of BCE botanical interventions remains a challenge, so does identifying the active principles, including any and all active components or metabolites as well as dosing concentrations responsible for the botanical's reported positive clinical outcomes. For those trials of BCE that have not shown benefit, such results may be a reflection of the wide variety of the clinical trial settings, BCE formulations, differences in their standardization schemes, and dosing strengths that researchers have used to investigate the clinical benefit of this botanical in assessing menopausal symptom relief. These four parameters alone represent a four-dimensional set of parameters, which in practice are really impossible to align when comparing different studies. While this alignment is theoretically possible, many studies and/or reports lack the level of detail to enable such comparison.

The third group of factors with major impact on clinical outcome relates to the basic pharmacology of the intervention materials. Research studies have employed a variety of BCE formulations, which have variously been described by all of the following terms: (a) "Black Cohosh" 20 mg/day standardized to 1 mg triterpene glycosides as 27-deoxyactein (*124*); (b) 6.5 mg of dry rhizome extract, from 60% ethanol extraction with a drug extract ratio: 4.5–8.5:1 (*125*, *126*); (c) "Black Cohosh" dried aqueous/ethanolic 58% *v/v* (*127*); (d) "Black Cohosh" standardized to 1 mg triterpene glycosides calculated as 27-deoxyactein, corresponding to 3.75 mg *Cimicifuga Rhizoma*[2] (*128*); (e) dried aqueous/ethanolic (58% *v/v*) extract of rhizome (*129*); (f) a 70% ethanol extract of roots and rhizomes (*119*). These are just examples and this list could be extended. Some studies reveal more about the formulation of their BCE interventions than others. However, no investigator has thus far been in a position to define *the* active principle(s) at work in those trials that have demonstrated that BCE relieves menopausal symptoms. Accordingly, a challenge remains to fill the gap between the several known *in vitro*- and *in vivo*-active principles in BCEs, the numerous recently discovered phytoconstituents of the plant that have not been studied biologically, the sophistication of standardization

---

[2] It should be noted that *Cimicifuga Rhizoma* is not produced from *A. racemosa.*

schemes required to capture this chemical diversity, and the complex nature of the clinical endpoints associated with menopausal symptom management. This again opens a four-dimensional set of complex parameters, which yield a hyper-complex correlation space for BCEs altogether.

Interestingly, clinical trial reporting guidelines began to emerge in the late 1990s, emphasizing increased transparency. The consolidated standards of reporting trials, or CONSORT guidelines, continue to be revised periodically, with a 2010 publication providing a checklist of 25 elements that should be included in parallel, randomized trial reports (*130*). *Gagnier et al.* have focused on specific recommendations on the reporting of clinical trials of herbal interventions (*131*) The elements emphasized by this group include the herbal intervention's Latin name, proprietary product name and country of registration, parts of the plant used, type of product (fresh or dry), solvent used, method of authentication, dosage, duration of administration, content (excipients), standardization, quantity of marker constituents, chemical fingerprint, any special testing, rationale for placebo and description of practitioners. This information complements the Product Integrity specifications for research funded by NIH's NCCAM (National Center for Complementary and Alternative Medicine; see NOT-AT-10-006 and at http://www.nccam.nih.gov/research/policies/naturalproduct.htm). A robust example of botanical intervention descriptions can be found in the UIC/NIH Botanical Center study of BCE and red clover used to ameliorate menopausal vasomotor symptoms (*116*). The information provided represents a transparent summary of a contiguous step-by-step process for developing a botanical dietary supplement, which begins with the acquisition of raw material and appropriate placement of associated voucher specimens, and extends to details of the final dosage form with comprehensive analytical testing information presented. Providing this information enables reviewers and investigators to focus on trial results, thereby obviating questions of product integrity that could challenge those results. It is also reflective of thoughtful advanced planning in a clinical trial design process that acknowledges the complexity of testing botanical interventions.

At the same time, it is important to note that concurrent progress in the chemical and pharmacological description of a botanical intervention material like BCE produces a constant paradigmatic shift with regard to the assignment of botanical marker compound(s) and botanical standardization. The case of BCE is particularly interesting, as the very recent discovery of several classes of previously unidentified alkaloids in a BCE has the potential to significantly change the integrated view of the clinical and *in vivo* potential and targets of this plant extract.

With the variety and scope of trial designs, BCE formulations, dosages and intervention outcome measures, as well as conflicting trial results claiming both effective as well as non-effective outcomes, it is not surprising that some BCE clinical trial reviewers have called for more extensive and rigorous clinical trials (*97*, *132*, *133*). Two meta-analyses of BCE trials have been published; the authors of each study separately suggest that as a group these trials are too heterogeneous, thus limiting the studies that could be grouped for evaluation by meta-analysis. *Shams et al.* in their meta-analysis of nine randomized clinical trials found that BCE improved vasomotor symptoms by 26% (*134*).

In contrast, *Leach et al.* from the Cochrane Collaboration examined 16 randomized clinical trials, finding no significant difference in the frequency of hot flashes between BCE treated and placebo groups, adding that the quality of the studies examined is uncertain enough that improved reporting methods would be highly beneficial (*135*). The authors not only conclude that more research on BCE safety and efficacy is warranted, but also that there is adequate justification for conducting further trials. Considering the recent findings of new chemical diversity in BCEs (see Sect. 2.3), it is likely that new active principles in the plant are yet to be discovered, and that phytochemical and other basic research has significant potential to widen the perspective of clinical research on this widely used botanical.

Additionally, meeting the challenge of devising clinical trial designs that adequately capture the impact of botanical interventions on human health is vitally important. Research in this area continues at an exciting time, in which the study of human genomics and drug and/or botanical interactions is at its nascence. Investigators may find that botanicals like some ethical drugs produce a wide continuum of responses in various populations, reinforcing the need to investigate the pharmacokinetic and pharmacodynamic activities of both known and newly recognized *A. racemosa* chemical constituents. In accordance with the views of so many investigators, the BCE clinical research horizon must be broadened as researchers continue to understand this botanical in light of its long history of use, which continues into the twenty-first century.

7   Concluding Remarks

The last 30 years of research of *A. racemosa* as a medicinal plant or an herbal dietary supplement have produced much valuable knowledge of its botany, chemistry, and biology. Particularly, the ten-year research project led by the University of Illinois at Chicago (UIC)/National Institutes of Health (NIH) Botanical Center has for the first time carried out systematic and comprehensive studies of this plant in a unified interdisciplinary setting and significantly improved the understanding and interrelationship of its phytochemical constituents and pharmacological activities.

Recent efforts have revealed a much more detailed profile of the characteristic secondary metabolites in *A. racemosa*, which include almost 50 known cycloartane triterpenes, 10 cimicifugic acids, and more than 70 nitrogen-containing constituents. This information is not only useful in the fingerprint profiling for the botanical identification and authentication, but it also provides the chemical evidence and inspiration for the discovery of the pharmacological activities of this plant. The use of advanced spectroscopic techniques, such as MS and NMR, and computational approaches enables the rapid and reliable identification of these compounds directly from their mixtures (*8*, *35*), making them a great resource for the future studies on the phytoconstituents and quality standardization of *A. racemosa* and its related botanical preparations. Furthermore, new nomenclature has been established to

standardize the naming systems for the *Actaea* cycloartane triterpenes and cimici-
fugic acids, both of which are comprised of a number of congeneric compounds
(*8*, *24*) These naming systems provide scientists with some basic knowledge of
structural characteristics of the compounds concerned.

A. racemosa has a well-established traditional use for the alleviation of women's
menopausal symptoms such as hot flashes, insomnia, and depression. This has been
interpreted frequently as being due to the presence of phytoestrogens in *A. race-
mosa*, which would contribute to an estrogenic activity of the plant. However, phar-
macological evaluation of either the crude extracts or isolated phytoconstituents has
not yet provided the required evidence for this mode of action, but has instead
pointed to the lack of this type of pharmacological activity for BCEs.

Stress-related conditions, such as anxiety and depression, are among the major
symptoms at the menopausal stage. The ability of *A. racemosa* to attenuate the psy-
chological and physiological stress has been evaluated extensively and confirmed in
animals. This activity, at least on the *in vitro* level, is found to be partly ascribed to
the CNS effects of the alkaloid N_ω-methylserotonin, which has potent 5-HT$_7$ recep-
tor binding ability (*33*). In addition, four cycloartane triterpenes, including actein
(**24**), 23-*O*-acetylshengmanol-3-*O*-β-D-xylopyranoside (**32**), cimigenol-3-*O*-β-D-
xylopyranoside (**1**), and 25-*O*-acetylcimigenol-3-*O*-α-L-arabinopyranoside (**6**), are
also found to be CNS active, showing GABAergic activity (*11*). Other studies have
found *A. racemosa* extract effectively counteracts the metaphyseal bone loss of
the femur in control untreated ovariectomized animals (*85*). Further investigations have
ascribed this bioactivity to two cycloartane triterpenes, namely, 25-*O*-acetylcimigenol-
3-*O*-β-D-xylopyranoside (**5**) and 23-*epi*-26-deoxyactein (**29**) (*86*, *87*). The prelimi-
nary studies on the anticancer activity of *A. racemosa* have demonstrated that both
cimicifugic acids and cycloartane triterpenes are effective in inhibiting the growth
of breast and liver cancer cells (*7*, *90*, *91*). Hepatotoxicity has been the only safety
concern on *A. racemosa* consumption, and this topic has been considered in detail.
Thus, the vast majority of cases of reported acute hepatitis did not stand up to fur-
ther investigation, and none met the rigorous requirements of causality. Moreover,
in clinical studies, there has been no convincing evidence that *A. racemosa* causes
hepatotoxicity while the possibility exists of rare idiosyncratic cases. This has also
raised the concern about possible adulteration of *A. racemosa* either intentionally or
accidentally, especially by related species, some of which are similar superficially.

As seen in previous studies, the pharmacological evaluation was mainly concen-
trated on the individual compounds. Research in this way continues as long as other
new compounds are identified. The empirical knowledge of complementary and
alternative medicine suggests the therapeutic effects of herbal remedies possibly
result from the synergistic or additive interactions of multiple components (*136*).
Based on this concept, it can be hypothesized that this situation may also exist in
A. racemosa. Therefore, in future studies, it is absolutely necessary to explore the
potential synergism of the phytoconstituents of *A. racemosa*. To ascribe the chemi-
cal components to this mode of action, it requires a detailed profile of phytochemis-
try to furnish both qualitative and quantitative information, which can be obtained
only from large-scale metabolomic profiling of the plant. As a result of the chemical

complexity, and of current knowledge of this profile, the process will be challenging but possible by the use of advanced chromatographic and spectroscopic techniques as well as sophisticated data mining tools (*137*). Establishment of such a metabolomic profile will provide a deeper understanding of the secondary metabolism in *A. racemosa* and help map the biosynthesis pathways, identification and understanding of the active secondary metabolites.

Although almost 50 clinical studies of the effects of BCEs on menopausal symptoms have been reported, only a small minority of these have been randomized double-blinded controlled studies and virtually all of those have been underpowered. Attempts at meta-analyses of these have been complicated by the use of different preparations, and result parameters. Investigators are unanimous in only one aspect: the need for larger more rigorously controlled studies (*97*, *132–135*).

In conclusion, while recent advances in basic (phyto)chemical and biological as well as clinical knowledge of *A. racemosa* have not been able to provide the ultimate rationale for the safe and efficacious use of this popular botanical, there is an increasingly solid foundation to justify further exploration of the metabolomic diversity and biological impact of *A. racemosa* phytoconstituents, and to conduct clinical trials with rigorously defined intervention materials, possibly targeting previously underexplored clinical endpoints.

References

1. Moermann DE (1986) Medicinal Plants of Native America. Museum of Anthropology, University of Michigan, Ann Arbor, Michigan, USA
2. Anonymous 1820, The Pharmacopœia of the United States of America (the United States Pharmacopœia) 1820-19, Easton, Pennsylvania, USA
3. Blumenthal M, Lindstrom A, Ooyen C, Lynch ME (2012) Herb Supplement Sales Increase 4.5% in 2011. HerbalGram 95:60
4. Foster S (2013) Exploring the Peripatetic Maze of Black Cohosh Adulteration - a Review of the Nomenclature, Distribution, Chemistry, Market Status, Analytical Methods and Safety Concerns of This Popular Herb. HerbalGram 98:32
5. Li JX, Yu ZY (2006) *Cimicifugae Rhizoma*: From Origins, Bioactive Constituents to Clinical Outcomes. Curr Med Chem 13:2927
6. Imai A, Yao P, Overk C, Bolton JL, Nikolic D, van Breemen RB, Pauli GF (2008) Phytochemical Investigation of the Estrogenic Aerial Parts of *Cimicifuga racemosa* (L.) Nutt. (Ranunculaceae), 2008 Interim ASP and Oxford ISCB Meeting, Oxford, Mississippi, USA
7. Mimaki Y, Nadaoka I, Yasne M, Ohtake Y, Ikeda M, Watanabe K, Sashida Y (2006) Neocimicigenosides A and B, Cycloartane Glycosides from the Rhizomes of *Cimicifuga racemosa* and Their Effects on CRF-Stimulated ACTH Secretion from AtT-20 Cells. J Nat Prod 69:829
8. Qiu F, Imai A, McAlpine J, Lankin DC, Burton IW, Karakach TK, Farnsworth NR, Pauli GF (2012) Dereplication, Residual Complexity and Rational Naming – the Case of the *Actaea* Triterpenes. J Nat Prod 75:432
9. Friesen JB, Pauli GF (2005) G.U.E.S.S. - a Generally Useful Estimate of Solvent Systems. J Liq Chromatogr Relat Technol 28:2877
10. Chen S-N, Fabricant DS, Lu Z-Z, Fong HHS, Farnsworth NR (2002) Cimiracemosides I-P, New 9,19-Cyclolanostane Triterpene Glycosides from *Cimicifuga racemosa*. J Nat Prod 65:1391

11. Cicek SS, Schwaiger S, Ellmerer E, Stuppner H (2010) Development of a Fast and Convenient Method for the Isolation of Triterpene Saponins from *Actaea racemosa* by High-Speed Countercurrent Chromatography Coupled with Evaporative Light Scattering Detection. Planta Med 76:467

12. Friesen JB, Pauli GF (2007) Rational Development of Solvent System Families in Counter-Current Chromatography. J Chromatogr A 1151:51

13. Zhou L, Yang JS, Tu GZ, Zou JH (2006) Cyclolanostane Triterpene Glycosides from *Souliea vaginata*. Chem Pharm Bull 54:823

14. Chen S-N, Lankin DC, Nikolic D, Fabricant DS, Lu ZZ, Ramirez B, van Breemen RB, Fong HHS, Farnsworth NR, Pauli GF (2007) Chlorination Diversifies *Cimicifuga racemosa* Triterpene Glycosides. J Nat Prod 70:1016

15. Napolitano J, Lankin DC, Graf TN, Friesen JB, Chen S-N, McAlpine J, Oberlies NH, Pauli GF (2013) HiFSA Fingerprinting of Isomers with Near Identical NMR Spectra: The Silybin/Isosilybin Case. J Org Chem 78:2827

16. Napolitano JG, Gödecke T, Rodriguez Brasco MF, Jaki BU, Chen S-N, Lankin DC, Pauli GF (2012) The Tandem of Full Spin Analysis and QHNMR for the Quality Control of Botanicals Exemplified with *Ginkgo biloba*. J Nat Prod 75:238

17. Inokuma Y, Yoshioka S, Ariyoshi J, Arai T, Hitora Y, Takada K, Matsunaga S, Rissanen K, Fujita M (2013) X-Ray Analysis on the Nanogram to Microgram Scale Using Porous Complexes. Nature 495:461

18. Chen S-N, Li W, Fabricant DS, Santarsiero BD, Mesecar AD, Fitzloff JF, Fong HHS, Farnsworth NR (2002) Isolation, Structure Elucidation, and Absolute Configuration of 26-Deoxyactein from *Cimicifuga racemosa* and Clarification of Nomenclature Associated with 27-Deoxyactein. J Nat Prod 65:601

19. He K, Zheng BL, Kim CH, Rogers L, Zheng QY (2000) Direct Analysis and Identification of Triterpene Glycosides by LC/MS in Black Cohosh, *Cimicifuga racemosa*, and in Several Commercially Available Black Cohosh Products. Planta Med 66:635

20. Avula B, Wang YH, Smillie TJ, Khan IA (2009) Quantitative Determination of Triterpenoids and Formononetin in Rhizomes of Black Cohosh (*Actaea racemosa*) and Dietary Supplements by Using UPLC-UV/ELS Detection and Identification by UPLC-MS. Planta Med 75:381

21. Gödecke T, Napolitano JG, Rodriguez Brasco MF, Chen S-N, Jaki BU, Lankin DC, Pauli GF (2013) Validation of a Generic QHNMR Method for Natural Products Analysis. Phytochem Anal 24:581

22. Pauli GF, Jaki BU, Gödecke T, Lankin DC (2012) Quantitative ^{1}H NMR: Development and Potential of a Method for Natural Products Analysis - an Update. J Nat Prod 75:834

23. Pauli GF, Jaki BU, Lankin DC (2005) Quantitative ^{1}H NNMR: Development and Potential of a Method for Natural Products Analysis. J Nat Prod 68:133

24. Gödecke T, Nikolic D, Lankin DC, Chen S-N, Powell SL, Dietz B, Bolton JL, van Breemen RB, Farnsworth NR, Pauli GF (2009) Phytochemistry of Cimicifugic Acids and Associated Bases in Black Cohosh Root Extracts. Phytochem Anal 20:120

25. Kruse SO, Löhning A, Pauli GF, Winterhoff H, Nahrstedt A (1999) Fukiic and Piscidic Acid Esters from the Rhizome of *Cimicifuga racemosa* and the Estrogenic Activity of Fukinolic Acid. Planta Med 65:763

26. Finnemore H (1909) Constituents of Rhizome of *Cimicifuga racemosa*. Pharm J 83:145

27. Gemeinhardt K, Sade H, Schenck G (1956) Beiträge zu Einem Ergänzungsbuch für das Homöopathische Arzneibuch. 3. Mitteilung: *Cimicifuga racemosa* (L.) Nutt. Pharm Z 101:238

28. Swain T (1966) Comparative Phytochemistry. Academic Press, London, New York, p xiii

29. Laidlaw PP (1912) Laburnum Poisoning and Cytisine. Proc Royal Soc Med 5:10

30. Crum JD, Cassady JM, Olmstead PM, Picha NJ (1965) The Chemistry of Alkaloids. I. The Screening of Some Native Ohio Plants. Proc West Virginia Acad Sci 37:143

31. Furgiuele AR, Farnsworth NR, Buckley JP (1962) False-Positive Alkaloid Reactions Obtained with Extracts of *Piper methysticum*. J Pharm Sci 51:1156

32. Fabricant D, Nikolic D, Lankin DC, Chen S-N, Jaki B, Krunic A, van Breemen RB, Fong HHS, Farnsworth NR, Pauli GF (2005) Cimipronidine: A Novel Cyclic Guanidine Alkaloid from *Cimicifuga racemosa* (L.) Nutt. J Nat Prod 68:1266

33. Powell SL, Gödecke T, Nikolic D, Chen S-N, Dietz B, Farnsworth NR, van Breemen RB, Lankin DC, Pauli GF, Bolton JL (2008) *In Vitro* Serotonergic Activity of Black Cohosh and Identification of N_ω-Methylserotonin as Active Constituent. J Agric Food Chem 56:11718

34. Gödecke T, Lankin DC, Nikolic D, Chen S-N, van Breemen RB, Farnsworth NR, Pauli GF (2009) Guanidine Alkaloids and *Pictet-Spengler* Adducts from Black Cohosh (*Cimicifuga racemosa*). J Nat Prod 72:433

35. Nikolić D, Gödecke T, Chen S-N, White J, Lankin DC, Pauli GF, van Breemen RB (2012) Mass Spectrometric Dereplication of Nitrogen-Containing Constituents of Black Cohosh (*Cimicifuga racemosa* L.). Fitoterapia 83:441

36. Stevigny C, Bailly C, Quetin-Leclercq J (2005) Cytotoxic and Antitumor Potentialities of Aporphinoid Alkaloids. Curr Med Chem Anticancer Agents 5:173

37. Wood KV, Bonham CC, Miles D, Rothwell AP, Peel G, Wood BC, Rhodes D (2002) Characterization of Betaines Using Electrospray MS/MS. Phytochemistry 59:759

38. Morin AM (1984) Beta-Carboline Kindling of the Benzodiazepine Receptor. Brain Res 321:151

39. Lippke KP, Schunack WG, Wenning W, Muller WE (1983) Beta-Carbolines as Benzodiazepine Receptor Ligands. 1. Synthesis and Benzodiazepine Receptor Interaction of Esters of Beta-Carboline-3-Carboxylic Acid. J Med Chem 26:499

40. Hagen TJ, Skolnick P, Cook JM (1987) Synthesis of 6-Substituted Beta-Carbolines that Behave as Benzodiazepine Receptor Antagonists or Inverse Agonists. J Med Chem 30:750

41. Ishida J, Wang HK, Bastow KF, Hu CQ, Lee KH (1999) Antitumor Agents 201. Cytotoxicity of Harmine and Beta-Carboline Analogs. Bioorg Med Chem Lett 9:3319

42. Xu Z, Chang FR, Wang HK, Kashiwada Y, McPhail AT, Bastow KF, Tachibana Y, Cosentino M, Lee KH (2000) Anti-HIV Agents 45 (1) and Antitumor Agents 205 (2). Two New Sesquiterpenes, Leitneridanins A and B, and the Cytotoxic and Anti-HIV Principles from *Leitneria floridana*. J Nat Prod 63:1712

43. Tsuda M, Watanabe D, Kobayashi J (1998) Maeganedin A, a New Manzamine Alkaloid from *Amphimedon* Sponge. Tetrahedron Lett 39:1207

44. Di Giorgio C, Delmas F, Ollivier E, Elias R, Balansard G, Timon-David P (2004) *In Vitro* Activity of the Beta-Carboline Alkaloids Harmane, Harmine, and Harmaline toward Parasites of the Species *Leishmania infantum*. Exp Parasitol 106:67

45. Da Cunha EV, Fechinei IM, Guedes DN, Barbosa-Filho JM, Da Silva MS (2005) Protoberberine Alkaloids. Alkaloids Chem Biol 62:1

46. Grycova L, Dostal J, Marek R (2007) Quaternary Protoberberine Alkaloids. Phytochemistry 68:150

47. Schmidt J, Boettcher C, Kuhnt C, Kutchan TM, Zenk MH (2007) Poppy Alkaloid Profiling by Electrospray Tandem Mass Spectrometry and Electrospray FT-ICR Mass Spectrometry after [Ring-^{13}C6]-Tyramine Feeding. Phytochemistry 68:189

48. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR (2007) Proposed Minimum Reporting Standards for Chemical Analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211

49. Dookeran NN, Yalcin T, Harrison AG (1996) Fragmentation Reactions of Protonated α-Amino Acids. J Mass Spectrom 31:500

50. Kuhnert N, Jaiswal R, Matei MF, Sovdat T, Deshpande S (2010) How to Distinguish between Feruloyl Quinic Acids and Isoferuloyl Quinic Acids by Liquid Chromatography/Tandem Mass Spectrometry. Rapid Commun Mass Spectrom 24:1575

51. Somei M, Teranishi S, Yamada K, Yamada F (2001) The Chemistry of Indoles. CVII. A Novel Synthesis of 3,4,5,6-Tetrahydro-7-hydroxy-1*H*-azepino[5,4,3-*cd*]indoles and a New Finding on *Pictet-Spengler* Reaction. Chem Pharm Bull 49:1159

52. Facchini PJ (2001) Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. Ann Rev Plant Physiol Plant Mol Biol 52:29

53. Stevigny C, Jiwan JL, Rozenberg R, de Hoffmann E, Quetin-Leclercq J (2004) Key Fragmentation Patterns of Aporphine Alkaloids by Electrospray Ionization with Multistage Mass Spectrometry. Rapid Commun Mass Spectrom 18:523

54. Zhang Y, Shi Q, Shi P, Zhang W, Cheng Y (2006) Characterization of Isoquinoline Alkaloids, Diterpenoids and Steroids in the Chinese Herb *Jin-Guo-Lan* (*Tinospora sagittata* and *Tinospora capillipes*) by High-Performance Liquid Chromatography/Electrospray Ionization with Multistage Mass Spectrometry. Rapid Commun Mass Spectrom 20:2328

55. Henion JD, Mordehai AV, Cai J (1994) Quantitative Capillary Electrophoresis-Ion Spray Mass Spectrometry on a Benchtop Ion Trap for the Determination of Isoquinoline Alkaloids. Anal Chem 66:2103

56. Guo Y, Kojima K, Lin L, Fu X, Zhao C, Hatano K, Chen Y-J, Ogihara Y (1999) A New N-Methyltetrahydroprotoberberine Alkaloid from *Tinospora hainanesis*. Chem Pharm Bull 47:287

57. Matsuda H, Suzuki Y (1984) γ-Guanidinobutyraldehyde Dehydrogenase of *Vicia faba* Leaves. Plant Physiol 76:654

58. Musshoff F, Lachenmeier DW, Schmidt P, Dettmeyer R, Madea B (2005) Systematic Regional Study of Dopamine, Norsalsolinol, and (*R/S*)-Salsolinol Levels in Human Brain Areas of Alcoholics. Alcohol Clin Exp Res 29:46

59. Reuter G, Diehl HJ (1971) Guanidine Derivatives in *Leonurus sibiricus* L. Pharmazie 26:777

60. Atta-ur-Rahman, Bhatti MK, Akhtar F, Choudhary MI (1992) Alkaloids of *Fumaria indica*. Phytochemistry 31:2869

61. Verbitski SM, Gourdin GT, Ikenouye LM, McChesney JD, Hildreth J (2008) Detection of *Actaea racemosa* Adulteration by Thin-Layer Chromatography and Combined Thin-Layer Chromatography-Bioluminescence. J AOAC Int 91:268

62. He K, Pauli GF, Zheng B, Wang H, Bai N, Peng T, Roller M, Zheng Q (2006) *Cimicifuga* Species Identification by High Performance Liquid Chromatography-Photodiode Array/ Mass Spectrometric/Evaporative Light Scattering Detection for Quality Control of Black Cohosh Products. J Chromatogr A 1112:241

63. Jiang B, Ma C, Motley T, Kronenberg F, Kennelly EJ (2011) Phytochemical Fingerprinting to Thwart Black Cohosh Adulteration: A 15 *Actaea* Species Analysis. Phytochem Anal 22:339

64. Li W, Chen S, Fabricant D, Angerhofer CK, Fong HHS, Farnsworth NR, Fitzloff JF (2002) High-Performance Liquid Chromatographic Analysis of Black Cohosh (*Cimicifuga racemosa*) Constituents with in-Line Evaporative Light Scattering and Photodiode Array Detection. Anal Chim Acta 471:61

65. Jiang B, Kronenberg F, Nuntanakorn P, Qiu M-H, Kennelly EJ (2006) Evaluation of the Botanical Authenticity and Phytochemical Profile of Black Cohosh Products by High-Performance Liquid Chromatography with Selected Ion Monitoring Liquid Chromatography-Mass Spectrometry. J Agric Food Chem 54:3242

66. Avula B, Ali Z, Khan IA (2007) Chemical Fingerprinting of *Actaea racemosa* (Black Cohosh) and Its Comparison Study with Closely Related *Actaea* Species (*A. pachypoda, A. podocarpa, A. rubra*) by HPLC. Chromatographia 66:757

67. Jiang B, Kronenberg F, Balick MJ, Kennelly EJ (2006) Analysis of Formononetin from Black Cohosh (*Actaea racemosa*). Phytomedicine 13:477

68. Li W, Sun Y, Liang W, Fitzloff JF, van Breemen RB (2003) Identification of Caffeic Acid Derivatives in *Actaea racemosa* (*Cimicifuga racemosa*, Black Cohosh) by Liquid Chromatography/Tandem Mass Spectrometry. Rapid Commun Mass Spectrom 17:978

69. Zerega N, Mori S, Lindqvist C, Zheng Q, Motley T (2002) Using Amplified Fragment Length Polymorphisms (AFLP) to Identify Black Cohosh (*Actaea racemosa*). Econ Bot 65:164

70. Motley TJ, Lück L, Zerega NJC (2004) Genetic Diversity and DNA Fingerprinting of Black Cohosh (*Actaea racemosa*). Proc Global Summit Med Pl 1:112

71. Baker DA, Stevenson DW, Little DP (2012) DNA Barcode Identification of Black Cohosh Herbal Dietary Supplements. J AOAC Int 95:1023

72. Compton JA, Culham A, Jury SL (1998) Reclassification of *Actaea* to Include *Cimicifuga* and *Souliea* (Ranunculaceae): Phylogeny Inferred from Morphology, NRDNA, ITS, and CPDAN TRN-F Sequence Variation. Taxon 47:593

73. Jamróz MK, Jamróz MH, Dobrowolski JC, Gliński JA, Gleńsk M (2012) One New and Six Known Triterpene Xylosides from *Cimicifuga racemosa*: FT-IR, Raman and NMR Studies and DFT Calculations. Spectrochimica Acta Part A 93:10

74. Jamróz MK, Janróz MH, Dobrowolski JC, Glinski JA, Davey MH, Wawer I (2011) Novel and Unusual Triterpene from Black Cohosh. Determination of Structure of 9,10-Seco-9,19-Cyclolanostane Xyloside (Cimipodocarpaside) by NMR, IR and Raman Spectroscopy and DFT Calculations. Spectrochimica Acta, Part A 78:107

75. Jiang Y, David B, Tu P, Barbin Y (2010) Recent Analytical Approaches in Quality Control of Traditional Chinese Medicines - a Review. Anal Chim Acta 657:9

76. Bedir E, Khan IA (2000) Cimiracemoside A: A New Cyclolanostanol Xyloside from the Rhizome of *Cimifuga racemosa*. Chem Pharm Bull 48:425

77. Shao Y, Harris A, Wang MF, Zhang HJ, Cordell GA, Bowman M, Lemmo E (2000) Triterpene Glycosides from *Cimicifuga racemosa*. J Nat Prod 63:905

78. Kondo Y, Takemoto T (1972) The Structure of Cimifugin, a New Bitter Principle from *Cimicifuga simplex* Wormsk. Chem Pharm Bull 20:1940

79. Burdette JE, Liu J, Chen S-N, Fabricant DS, Piersen CE, Barker EL, Pezzuto JM, Mesecar AD, van Breemen RB, Farnsworth NR, Bolton JL (2003) Black Cohosh Acts as a Mixed Competitive Ligand and Partial Agonist of the Serotonin Receptor. J Agric Food Chem 51:5661

80. Liu JH, Burdette JE, Xu HY, Gu CG, van Breemen RB, Bhat KPL, Booth N, Constantinou AI, Pezzuto JM, Fong HHS, Farnsworth NR, Bolton JL (2001) Evaluation of Estrogenic Activity of Plant Extracts for the Potential Treatment of Menopausal Symptoms. J Agric Food Chem 49:2472

81. Einbond LS, Shimizu M, Ma H, Wu HA, Goldsberry S, Sicular S, Pnjikaran M, Genovese G, Cruz E (2008) Actein Inhibits the Na^+K^+-ATPase and Enhances the Growth Inhibitory Effect of Digitoxin on Human Breast Cancer Cells. Biochem Biophys Res Commun 375:608

82. Jarry H, Harnischfeger G, Düker E-M (1985) Studies on the Endocrine Efficacy of the Constituents of *Cimicifuga racemosa*: 2. *In Vitro* Binding of Constituents to Estrogen Receptors. Planta Med 51:316

83. Kennelly EJ, Baggett S, Nuntanakorn P, Ososki AL, Mori SA, Duke J, Coleton M, Kronenberg F (2002) Analysis of Thirteen Populations of Black Cohosh for Formononetin. Phytomedicine 9:4161

84. Panossian A, Danielyan A, Mamikonyan G, Wikman G (2004) Methods of Phytochemical Standardisation of Rhizoma *Cimicifugae Racemosae*. Phytochem Anal 15:100

85. Seidlová-Wuttke D, Hesse O, Jarry H, Christoffel V, Spengler B, Becker T, Wuttke W (2003) Evidence for Selective Estrogen Receptor Modulator Activity in a Black Cohosh (*Cimicifuga racemosa*) Extract: Comparison with Estradiol-17β. Eur J Endocrinol 149:351

86. Qiu SX, Dan C, Ding LS, Peng S, Chen S-N, Farnsworth NR, Nolta J, Gross ML, Zhou P (2007) A Triterpene Glycoside from Black Cohosh That Inhibits Osteoclastogenesis by Modulating RANKL and TNRalpha Signaling Pathways. Chem Biol 14:860

87. Ruhlen RL, Sun GY, Sauter ER (2008) Black Cohosh: Insights into Its Mechanism(s) of Action. Integr Med Insights 2008:21

88. Kim CD, Lee W-K, Lee M-H, Cho HS, Lee YK, Roh S-S (2004) Inhibition of Mast Cell-Dependent Allergy Reaction by Extract of Black Cohosh (*Cimicifuga racemosa*). Immunopharmacol Immunotoxicol 26:299

89. Lee JH, Cuong TD, Kwack SJ, Seok JH, Lee JK, Jeong JY, Woo M-H, Choi JS, Lee HK, Min BS (2012) Cycloartane-Type Triterpene Glycosides from the Rhizomes of *Cimicifuga heracleifolia* and Their Anticomplementary Activity. Planta Med 78:1391

90. Burdette JE, Chen S-N, Lu ZZ, Xu H, White BE, Fabricant DS, Liu J, Fong HHS, Farnsworth NR, Constantinou AI, van Breemen RB, Pezzuto JM, Bolton JL (2002) Black Cohosh

(*Cimicifuga racemosa* L.) Protects against Menadione-Induced DNA Damage through Scavenging of Reactive Oxygen Species: Bioassay-Directed Isolation and Characterization of Active Principles. J Agric Food Chem 50:7022

91. Einbond L, Saxe, Soffritti M, Esposti D, Degli, Park T, Cruz E, Su t, Wu H-A, Wang X, Zhang Y-J, Ham J, Goldberg IJ, Kronenberg F, Valdimirova A (2009) Actein Activates Stress- and Statin-Associated Responses and Is Bioavailable in Sprague-Dawley Rats. Fund Clin Pharmacol 23:311

92. Nadaoka I, Yasue M, Kitagawa Y, Koga Y (2012) Oral Administration of *Cimicifuga racemosa* Extract Attenuates Psychological and Physiological Stress Responses. Biomed Res 33:145

93. Nakazato K, Shimonaka Y (1982) The Japanese State-Trait Anxiety Inventory: Age and Sex Differences. Percept Motor Skills 69:611

94. Vitetta L, Thomsen M, Sali A (2003) Black Cohosh and Other Herbal Remedies Associated with Acute Hepatitis. Med J Aust 178:411

95. Lontos S, Jones RM, Angus PW, Gow PJ (2003) Acute Liver Failure Associated with the Use of Herbal Preparations Containing Black Cohosh. Med J Aust 179:390

96. Mahady GB, Low Dog T, Barrett ML, Chavez ML, Gardiner P, Ko R, Marles RJ, Pellicore LS, Giancaspro GI, Sarma DN (2008) United States Pharmacopeia Review of the Black Cohosh Case Reports of Hepatotoxicity. Menopause 15:628

97. Borrelli F, Ernst E (2008) Black Cohosh (*Cimicifuga racemosa*): A Systematic Review of Adverse Events. Am J Obstetr Gynecol 199:455

98. Roberts H (2010) Safety of Herbal Medicinal Products in Women with Breast Cancer. Maturitas 66:363

99. Teschke R, Schwarzenboeck A, Schmidt-Taenzer W, Wolff A, Hennermann KH (2011) Herb Induced Liver Injury Presumably Caused by Black Cohosh: A Survey of Initially Purported Cases and Herbal Quality Specifications. Ann Hepatol 10:249

100. Naser B, Schnitiker J, Minkin MJ, de Arriba SG, Noite K-U, Osmers R (2011) Suspected Black Cohosh Hepatotoxicity: No Evidence by Meta-Analysis of Randomized Controlled Clinical Trials for Isopropanolic Black Cohosh Extract. Menopause 18:366

101. Lüde S, Török M, Dieterle S, Knapp AC, Kaeufeler R, Jäggi R, Spornitz U, Krähenbühl S (2007) Hepatic Effects of *Cimicifuga racemosa* Extract *in Vivo* and *in Vitro*. Cell Mol Life Sci 64:2848

102. Mazzanti G, Di Sotto A, Franchitto A, Mastrangelo S, Pezzella M, Vitalone A, Mammola CL (2008) Effects of *Cimicifuga racemosa* Extracts on Liver Morphology and Hepatic Function Indices. Phytomedicine 15:1021

103. Campos LB, Gilglioni EH, Garcia R, Fernandes, Brito MdN, Natali MRM, Ishii-Iwamoto EL, Salgueiro-Pagadigorria CL (2012) *Cimicifuga racemosa* Impairs Fatty Acid Beta-Oxidation and Induces Oxidative Stress in Livers of Ovariectomized Rats with Renovascular Hypertension. Free Radical Biol Med 53:680

104. Fabricant DS, Farnsworth NR (2005) Black Cohosh (*Cimicifuga racemosa*), In: Coates P, Blackman M, Cragg G, Levine M, Moss J, White J (eds), Encyclopedia of Dietary Supplements. Marcel Dekker, New York, p. 41

105. Liske E (1998) Therapeutic Efficacy and Safety of *Cimicifuga racemosa* for Gynecologic Disorders. Adv Ther 15:45

106. Huntley A, Ernst E (2003) A Systematic Review of the Safety of Black Cohosh. Menopause 10:58

107. Low Dog T, Powell KL, Weisman SM (2003) Critical Evaluation of the Safety of *Cimicifuga racemosa* in Menopause Symptom Relief. Menopause 10:299

108. Huntley A (2004) The Safety of Black Cohosh (*Actaea racemosa, Cimicifuga racemosa*). Expert Opin Drug Safety 3:615

109. Cohen SM, O'Connor AM, Hart J, Merel NH, Te HS (2004) Autoimmune Hepatitis Associated with the Use of Black Cohosh: A Case Study. Menopause 11:575

110. Joy D, Joy J, Duane P (2008) Black Cohosh: A Cause of Abnormal Postmenopausal Liver Function Tests. Climacteric 11:84

111. Chow ECY, Teo M, Ring JA, Chen JW (2008) Liver Failure Associated with the Use of Black Cohosh for Menopausal Symptoms. Med J Aust 188:420

112. Teschke R, Schwarzenboeck A (2009) Suspected Hepatotoxicity by *Cimicifugae Racemosae* Rhizoma (Black Cohosh, Root): Critical Analysis and Structured Causality Assessment. Phytomedicine **16**:72

113. Teschke R, Bahre R, Genthner A, Fuchs J, Schmidt-Taenzer W, Wolff A (2009) Reply To: Suspected Black Cohosh Hepatotoxicity - Causality Assessment Versus Safety Signal. Quality Versus Quantity. Maturitas 64:141

114. Mahady G, Low Dog T, Sarma DN, Giancaspro GI (2009) Suspected Black Cohosh Hepatotoxicity - Causality Assessment Versus Safety Signal. Maturitas 64:139

115. Teschke R, Frenzel C, Glass X, Schulze J, Eickhoff A (2013) Herbal Hepatotoxicity: A Critical Review. Br J Clin Pharmacol 75:630

116. Geller SE, Shulman LP, van Breemen RB, Banuvar S, Epstein G, Hedayat S, Nikolic D, Krause EC, Piersen CE, Bolton JL, Pauli GF, Farnsworth NR (2009) Safety and Efficacy of Black Cohosh and Red Clover for the Management of Vasomotor Symptoms: A Randomized Controlled Trial. Menopause 16:1156

117. Newton KM, Reed SD, LaCroix AZ, Grothaus LC, Ehrlich K, Guiltinan J (2006) Treatment of Vasomotor Symptoms of Menopause with Black Cohosh, Multibotanicals, Soy, Hormone Therapy, or Placebo: A Randomized Trial. Ann Int Med 145:869

118. Carpenter JS, Newton KM, Sternfeld B, Joffe H, Reed SD, Ensrud KE, Milata JL (2012) Laboratory and Ambulatory Evaluation of Vasomotor Symptom Monitors from the Menopause Strategies Finding Lasting Answers for Symptoms and Health Network. Menopause 19:664

119. Chan BY, Lau KS, Jiang B, Kennelly EJ, Kronenberg F, Kung AWC (2008) Ethanolic Extract of *Actaea racemosa* (Black Cohosh) Potentiates Bone Nodule Formation in MC3T3-E1 Preosteoblast Cells. Bone 43:567

120. Li JX, Liu J, He CC, Yu ZY, Du Y, Kadota S, Seto H (2007) Triterpenoids from *Cimicifugae Rhizoma*, a Novel Class of Inhibitors on Bone Resorption and Ovariectomy-Induced Bone Loss. Maturitas 58:59

121. Borrelli F, Izzo AA, Ernst E (2003) Pharmacological Effects of *Cimicifuga racemosa*. Life Sci 73:1215

122. Rhyu MR, Lu J, Webster DE, Fabricant DS, Farnsworth NR, Wang ZJ (2006) Black Cohosh (*Actaea racemosa, Cimicifuga racemosa*) Behaves as a Mixed Competitive Ligand and Partial Agonist at the Human Mu Opiate Receptor. J Agric Food Chem 54:9852

123. Reame NE, Lukacs JL, Padmanabhan V, Eyvazzadeh AD, Smith Yolanda R, Zubieta J-K (2008) Black Cohosh Has Central Opioid Activity in Postmenopausal Women: Evidence from Naloxone Blockade and Positron Emission Tomography Neuroimaging. Menopause 15:832

124. Pockaj BA, Gallagher JG, Loprinzi CL, Stella PJ, Barton DL, Sloan JA, Lavasseur BI, Rao RM, Fitch TR, Rowland KM, Novotny PJ, Flynn PJ, Richelson E, Fauq AH (2006) Phase III Double-Blind, Randomized, Placebo-Controlled Crossover Trial of Black Cohosh in the Management of Hot Flashes: NCCYG Trial N01CC. J Clin Oncol 24:2836

125. Frei-Kleiner S, Schaffner W, Rahlfs VW, Bodmer C, Birkhauser M (2005) *Cimicifuga racemosa* Dried Ethanolic Extract in Menopausal Disorders: A Double-Blind Placebo-Controlled Clinical Trial. Maturitas 51:397

126. Drewe J, Zimmermann C, Zahner C (2013) The Effect of a *Cimicifuga racemosa* Extracts Ze 450 in the Treatment of Climacteric Complaints – an Observational Study. Phytomedicine 20:659

127. Raus K, Brucker C, Gorkow C, Wuttke W (2006) First-Time Proof of Endometrial Safety of the Special Black Cohosh Extract (*Actaea* or *Cimicifuga racemosa* Extract) CR BNO 1055. Menopause 13:678

128. Uebelhack R, Blohmer JU, Graubaum HJ, Busch R, Gruenwald J, Wernecke KD (2006) Black Cohosh and St. John's Wort for Climacteric Complaints - a Randomized Trial. Obstet and Gynecol 107:247

129. Wuttke W, Gorkow C, Seidlova-Wuttke D (2006) Effects of Black Cohosh (*Cimicifuga racemosa*) on Bone Turnover, Vaginal Mucosa, and Various Blood Parameters in Postmenopausal Women: A Double-Blind, Placebo-Controlled, and Conjugated Estrogens-Controlled Study. Menopause 13:185
130. Schulz KF, Altman DG, Moher D, Consort Group (2010) Consort 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. Ann Int Med 152:726
131. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C (2006) Reporting Randomized, Controlled Trials of Herbal Interventions: An Elaborated Consort Statement. Ann Int Med 144:364
132. Fabricant DS, Krause EC, Farnsworth NR (2010) Black Cohosh. In: Coates PM, Betz JM, Blackman MR (eds), Encyclopedia of Dietary Supplements. CRC Press, New York, p 60
133. Wuttke W, Jarry H, Haunschild J, Stecher G, Schuh M, Seidlova-Wuttke D (2014) The Non-Estrogenic Alternative for the Treatment of Climacteric Complaints: Black Cohosh (*Cimicifuga* or *Actaea racemosa*). J Ster Biochem Mol Biol 139:302
134. Shams T, Setia MS, Hemmings R, McCusker J, Sewitch M, Ciampi A (2010) Efficacy of Black Cohosh-Containing Preparations on Menopausal Symptoms: A Meta-Analysis. Alt Ther Health Med 16:36
135. Leach MJ, Moore V (2012) Black Cohosh (*Cimicifuga* spp.) for Menopausal Symptoms. Cochrane Database Systemat Rev 9:CD007244
136. Gertsch J (2011) Botanical Drugs, Synergy, and Network Pharmacology: Forth and Back to Intelligent Mixtures. Planta Med 77:1086
137. Sumner LW, Mendes P, Dixon RA (2003) Plant Metabolomics: Large-Scale Phytochemistry in the Functional Genomics Era. Phytochemistry 62:817
138. Watanabe K, Mimaki Y, Sakagami H, Sashida Y (2002) Cycloartane Glycosides from the Rhizomes of *Cimicifuga racemosa* and Their Cytotoxic Activities. Chem Pharm Bull 50:121
139. Linde H (1968) Die Inhaltsstoffe von *Cimicifuga racemosa*. 5. Mitt.: 27-Desoxyacetylacteol. Arch Pharm 301:335
140. Berger S, Junior P, Kopanski L (1988) 27-Deoxyactein: A New Polycyclic Triterpenoid Glycoside from *Actaea racemosa*. Planta Med 54:579
141. Linde H (1967) Die Inhaltsstoffe von *Cimicifuga racemosa*. 2. Mitt.: Zur Struktur des Acteins. Arch Pharm 300:885
142. Qiu F, Cai G, Jaki BU, Lankin DC, Franzblau SG, Pauli GF (2013) Quantitative Purity–Activity Relationships of Natural Products: The Case of Anti-Tuberculosis Active Triterpenes from *Oplopanax horridus*. J Nat Prod 76:413
143. Stöckigt J, Antonchick AP, Wu F, Waldmann H (2011) The *Pictet-Spengler* Reaction in Nature and in Organic Chemistry. Angew Chem Int Ed 50:8538
144. Kusano A, Takahira M, Shibano M, Miyase T, Okuyama T, Kusano G (1998) Studies on the constituents of *Cimicifuga* species. XXII. Structures of two new cyclolanostanol xylosides, cimiacerosides A and B. Heterocycles 48:1003

A Colorful History: The Evolution of Indigoids

Nicolas Gaboriaud-Kolar, Sangkil Nam, and *Alexios-Leandros Skaltsounis*

Contents

N. Gaboriaud-Kolar • A.-L. Skaltsounis (✉)
Department of Pharmacognosy and Natural Products Chemistry, School of Pharmacy,
University of Athens, Panepistimiopoulis Zografou, GR-15771 Athens, Greece
e-mail: ngaboriaud@pharm.uoa.gr; skaltsounis@pharm.uoa.gr

S. Nam
Molecular Medicine, Beckman Research Institute, City of Hope Comprehensive
Cancer Center, 1500 East Duarte Road, Duarte, CA 91010, USA
e-mail: snam@coh.org

1 Introduction

Indigoids are a fascinating class of natural products in which their chemical structures rely on the connection of two indole moieties. The most well-known member is probably indigotin, the main component of indigo dye used all over the world. However indigoids have found over the centuries several applications outside of dye chemistry and are now lead compounds in medicinal chemistry. The development of the indigoids has therefore led to a particular chemistry, characteristic of the several chemical fields in which they were applied.

1.1 Natural Dyes

Early in his history, mankind began manufacturing pigments for practical, esthetic, cosmetic or spiritual applications. The discovery of the oldest ochre-producing workshop dating from 100,000 years ago (Blomos Cave, South Africa (*1*)) illustrates this old know-how. However, the number of such dyes was limited by the naturally occurring possibilities and the first pigments were principally red, yellow blue or black, with the other colors being created by the mixing of these "primary colors". The chemical diversity of natural dyes varies from flavonoids (yellow dye), anthraquinones (red dye) through tannins (brown and black dyes) (*2*) to indigoids (blue and purple dyes). Their traditional sources could be divided in three categories: mineral, animal and vegetal. Mineral pigments are generally pure metals such as zinc (white), arsenic (green), or copper (green, blue, purple) at different oxidation states. They were used mainly for pictorial applications, and were of low interest for dyeing because of their instability or else because they were found to be toxic. Animal natural sources were also exploited such as insects as *Kermes vermilio* (cochineal). The Maya and Aztec civilizations from Central America produced the so-called *carmine* pigment, a red dye constituted of insoluble aluminum and calcium salts of carminic acid (**1**) (*3*). In the Mediterranean basin, the *Murex* shellfish family during antiquity provided the so-called "Tyrian purple", as worn by emperors. The deep purple color was given by indigotin and indirubin derivatives belonging to the indigoid family, which are the subject of this contribution.

1.2 Dyes from Plants

Above all, the vegetable kingdom represents the main source of sustainable natural pigments and knowledge of their extraction has been available for millennia. Nowadays, new natural pigments have emerged, although few have had a traditional
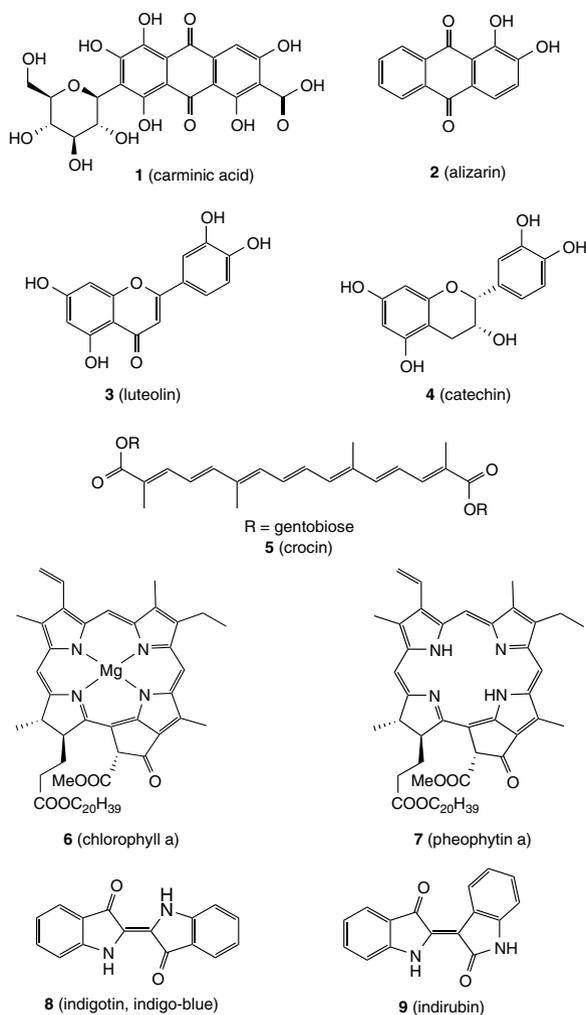
Fig. 1 Chemical structures of selected natural dyes

use throughout human history. The roots of *Rubia tinctoria* Salisb. (Rubiaceae) or madder were used by the Egyptians to produce the so-called "alizarin red" (*4*). This red pigment has been found in the *Tutankhamun* tomb and owes its color to the presence of the anthraquinone alizarin (**2**) (Fig. 1). *Reseda luteola* L. or weld (Brassicaceae) was used in France (thirteenth–seventeenth century) to dye yellow hats as a distinctive Jewish symbol. The yellow color was given by the flavonoid

luteolin (**3**) (*5*) (Fig. 1) first isolated by the French chemist *M. E. Chevreul*. The so-called "black-catechu" black dye was obtained in India from the tree bark and heartwood of *Acacia catechu* Willd. (Leguminosae). This complex mixture contains especially tannins, and catechin (**4**) (Fig. 1) is the most abundant representative compound. The stigmata of the flowers of *Crocus sativus* L. have been used for a long time notably in Persia. By crushing them, an orange-red pigment was extracted, for which the main chromophore is the carotenoid crocin (**5**) (Fig. 1). The crushed flowers were then applied directly on the cloth to afford an orange colored product. Surprisingly and although it is widely represented in Nature, a green color has not been successfully extracted from natural sources. The abundant chlorophyll (**6**) (Fig. 1) is not sufficiently heat-stable to be correctly and permanently bound to textile fibers. Moreover, it loses its coordinating magnesium under basic conditions (washing) to form pheophytin (**7**) (Fig. 1, olive-green color). The clothes were thus dyed in green using the toxic cupric oxide (mineral), explaining the few green-dye substances found.

The most widely used color in human history undoubtedly has been blue, including indigo extracted from indigo-bearing plants such as *Isatis* (Europe, Asia), *Indigofera* (India), or *Polygonum* (Japan, Asia) species. Around 300 species of plants in these genera exist, but only a few were exploited for dyeing, as described later in this contribution. Traces of indigo production have been found from England to South America and through to India. Indigotin (**8**) (Fig. 1), the main indigo constituent, belongs to the family of indigoids mentioned above and will be discussed further below. It is the only natural dye for which an industrial synthesis has been successfully realized, leading to its great success, especially in the blue jeans industry.

Until recently, the use of natural dyes was very limited due to the development of methods worked out for their chemical synthesis. However, new interest in natural products combined with the overall environmental awareness that has developed in the last ten years has led to their "rediscovery". Chlorophyll (**7**) is thus a useful tool known as an "eco-friendly" chemical to create more stable dyes (*6*). Indigo production took advantage of this "new" interest and local industries now produce it in an traditional way, as in the south of France (*7*). Nevertheless, its traditional production never stopped in India or in South America. Originally, artists such as *Inge Boesken Kanold* (*8*) found in Tyrian purple a new application as a high-class pigment in paintings. Indigo and Tyrian purple have symbolized the reputation of powerful people, made the fortunes of many merchants like the Phoenicians, and were a great source of benefits for European colonial empires. Their history has been deeply connected with the evolution of our society and they are still influential in our development. Moreover, the discovery of indirubin (**9**) (Fig. 1) as the main therapeutic compound of a traditional Chinese medicine (TCM) recipe has propelled this class of natural products from a dye to the field of medicinal chemistry.

2 Indigoid Family

2.1 *Family Presentation*

Indigoids are a natural class of organic compounds. They are bis-indole alkaloids connected by a double bond (of (*Z*) or (*E*) configuration) and can be substituted in their aromatic parts. "Indigo-blue" or indigotin (**8**) (Fig. 2) was the first representative of this family of natural products as the main chromophore of indigo dye. It is

8 (indigotin)
("indigo-bue")

10 (indigo)
(blue)

11 (isoindigo)
(brown)

9 (indirubin)
(reddish purple)

12 (isoindirubin)

13 (6-bromoindirubin: R^1 = Br ; R^2 = H)
14 (6'-bromoindirubin: R^1 = H ; R^2 = Br)
15 (6,6'-dibromoindirubin: R^1 = R^2 = Br)

16 (6-bromoindigotin: R^1 = Br ; R^2 = H)
17 (6,6'-dibromoindigotin: R^1 = R^2 = Br)

general structure of new indirubin derivatives
found in *H. trunculus*

18 (5,7,5',7'-tetrabromo-6,6'-dimethoxyindigotin)

19 (5,7,6'-tribromo-6-methoxyindigotin)

20 (akashin A: R^1 = R^2 = H)
21 (akashin B: R^1 = H, R^2 = COCH$_3$)

22 (akashin C: R^1, R^2 =)

Fig. 2 Members of the indigoid family

the symmetric dimer of two indole moieties connected by an (*E*)-double bond and exhibiting two internal hydrogen bonds. The (*Z*)-isomer (**10**) (Fig. 2) of indigotin (**8**) is also blue, but is found as a minor product during dye production (see Sect. 3.1.2, Chemistry). Isoindigo (**11**) (Fig. 2), the positional isomer of indigotin (**8**), is generally called "brown-indigo" due to its color in solution. Indirubin (**9**) (Fig. 2) is the second most important member of the indigoid group. It is an asymmetric compound and represents a combination of indigotin (**8**) and isoindigo (**11**) units. Isoindirubin (**12**) (Fig. 2), the (*E*)-isomer of indirubin, has also been identified in some indigo plants. The halogenated 6-bromo-, 6′-bromo- and 6,6′-bromoindirubins (**13–15**) (Fig. 2) were discovered in the Muricidae shellfish family (see Sect. 4.2.1) along with indigotin (**8**), indirubin (**9**), 6-bromoindigotin (**16**) and 6,6′-dibromoindigotin (**17**) (*9*). Recently, four new derivatives were identified in *Hexaplex trunculus* extracts using LC-MS/MS analysis: an imine group replaces the carbonyl at the 3′-position along with the presence of one or two bromine atoms (*10*). However, their unambiguous structures have not yet been elucidated. Two other brominated indigoids (**18** and **19**) (Fig. 2) have also been identified as constituents of the acorn worm, *Ptychodera flava laysanica* (*11*). Finally, the family of akashins A-C (**20–22**) (Fig. 2), the only example of naturally chlorinated indigoids, has been identified in a strain of the terrestrial *Streptomyces* sp. 48/1497 (*12*).

All members of the indigoid family are colored compounds. According to the type of isomer ((*Z*) or (*E*)), the symmetry and the degree of substitution, the resultant color ranges from blue (indigo) to red (indirubin) or purple for the marine brominated indigoids. Small structural modifications thus lead to a large range of colors all along the visible spectrum. Therefore, a simple question may be asked: what is the origin of the color?

2.2 Particular Electronic Effects: Origin of the Color

The origin of the color of indigotin (**8**) is a question that has stimulated scientific debate for more than a century, and a number of studies have been published on this subject. Summarized here are some of the most explicit ones.

The color of a molecule in solution or in solid phase is due to its ability to absorb light energy characterized by the maximum absorption wavelength (λ_{max}) on raising the ground energy state to an excited state. The *Plank* relation (Eq. (1)) links the energy difference (ΔE) and the wavelength. As a consequence, ΔE decreases as the λ_{max} increases.

$$\Delta E = hc / \lambda_{max} \tag{1}$$

In the solid-phase state, indigotin (**8**) possesses a dark-blue color and values of λ_{max} for amorphous and crystalline indigotin (**8**) have been measured at 640 and 680 nm, respectively. *Weintsein* and *Wyman* (*13*) have interpreted this large bathochromic effect with the formation of inter-molecular hydrogen bonds to form indigotin dimers (**23**) (or polymers), hence explaining the blue-dark aspect of solid
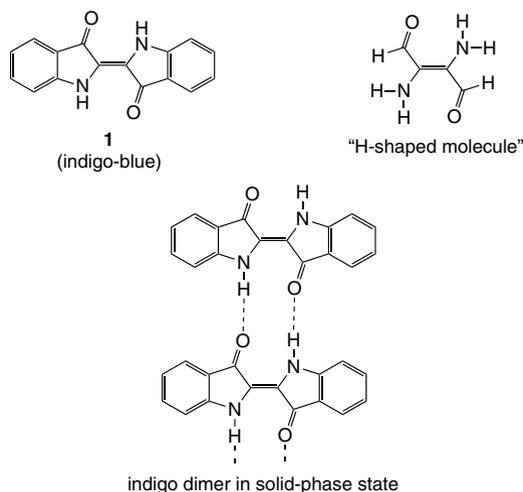
Fig. 3 Indigo dye and sources of color

indigotin (**8**). Primary rationalization studies led later by *Lüttke* and his collabora-
tors (*14*) established the chromophoric unit of indigo as the so-called "H-shaped
molecule" (**24**) (Fig. 3). This unit is localized on the hydrogen-bond area between
the two indole nitrogens and the two oxygen atoms, confirming the hypothesis of
dimerization through this region of the molecule.

Nonetheless, it has been shown that indigotin (**8**) maintains a strong "merocya-
nine" effect in solution. This is indicated by a strong bathochromic effect when the
dielectric constant of the solvent is increased. In this case, the bathochromic effect
is due to formation of hydrogen bonds with the solvent (*15*). Considering a polar
solvent (MeOH/H$_2$O/trifluoroacetic acid), the λ_{max} values of indigotin (**8**), indirubin
(**9**), and isoindigo (**11**) are 603, 552, and 500 nm (*16*). Referring to molecular orbital
interactions, the excitation related to λ_{max} corresponds to an HOMO→LUMO tran-
sition and moreover to a $\pi \to \pi^*$ transition: the HOMO lies on the nitrogen atoms
and the central double bond (ground state) whereas the LUMO lies mainly on the
simple C-C bond and oxygen atoms (excited state) (*17*, *18*). Indigoids are dyes of
the donor/acceptor type: the two indole nitrogen atoms are electron donors while
the two carbonyls groups are acceptors. Excitation due to light absorption delocal-
izes electrons through the bis-indole system. The representation of the resonance
structures (*17*, *18*) depicted in Schemes 1 and 2 illustrates the electron displace-
ments and explains the differences of absorption. For indigotin (**8**) (Scheme 1),
resonance forms lead to an extremely stable quadrupole structure: both the negative
and positive charges are stabilized on both sides of the double bond. The energy
difference between the ground state and the excited state is then small. Referring to
the *Planck* equation (Eq. (1)), this implies a long wavelength (603 nm).

For indirubin (**9**), the resonance forms lead to another quadrupole structure
(Scheme 2), which is less stabilized than the indigo form due to asymmetry.
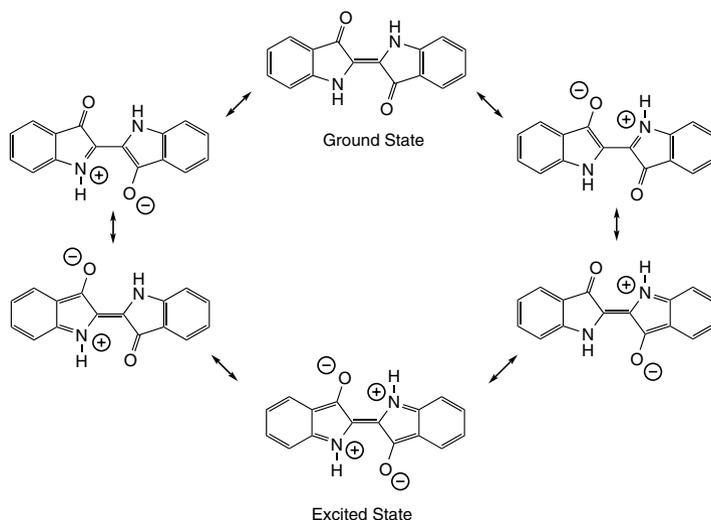
Scheme 1 Resonance forms of indigo
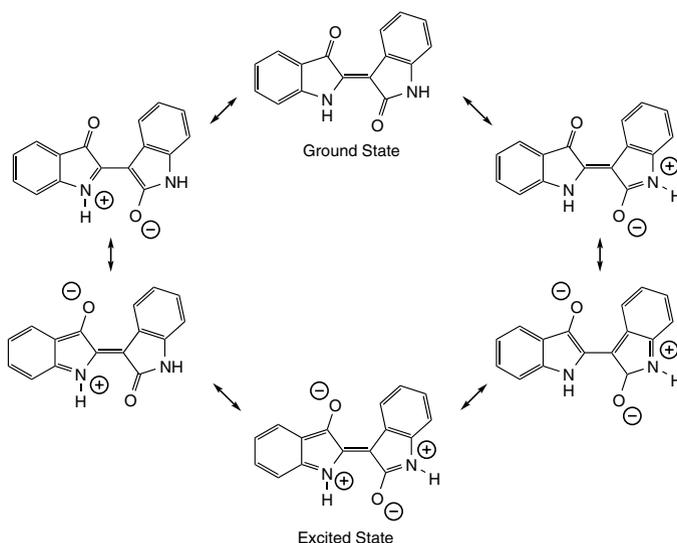
Scheme 2 Resonance forms of indirubin

However, the resonance structure of indirubin cannot be considered as a proof and further explorations have been undertaken.

The charge distribution from the resonance structures of indirubin and indigo were confirmed by a computational calculation approach based on a molecular orbital interaction (*17*). While the charge transfer is equally distributed on both sides of the indigotin core (**8**), it is localized mainly on the left portion ("indigo part")
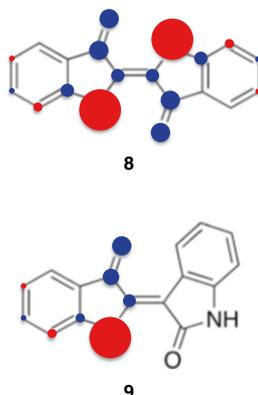
Fig. 4 Charge transfer in indigotin (**8**) and indirubin (**9**). (Reproduction from reference (*18*) prepared by the authors.)

of indirubin (*18*) (**9**) (Fig. 4), destabilizing the charge equilibrium. This destabilization leads to a bigger ΔE and so to a lower absorption length of 552 nm for indirubin (**9**), in the red area of the visible spectrum.

Indigoids possess a wide color range, resulting in high bathochromic effects due to their specific crystal network or solvent interactions. Those non-common electronic features still fascinate scientists and the use of these particular electronic effects is being applied for the development of new green organic electronics applications (*19*, *20*).

3 Indigo and Its Relatives

Among all the natural dyes known and extracted by human beings, indigo is one of the oldest (*21*). Indigo stripes were found on the linen of the clothes of mummies from the 3rd millennium BCE (Before the Common Era) (*22*). Pre-Columbian buildings were decorated by Maya people with paintings produced around 800 BCE, with the so-called Maya-Blue being a pigment constituted of mineral oxides and indigo (*23*). A cuneiform tablet from the Ancient Babylonian times (~600 BCE), gives the first written instructions for wool dyeing and represents the earliest extraction procedure for indigo (*22*). The use of indigo spread worldwide, but no-one these days is able to explain where and when all this great history started.

The etymologic origin of the word *indigo* came from the ancient Greek ινδικόν (pronounced "indikon"), which was translated to *indicum* in Latin and means "something from India". This supposes that the first contact with the pigment in South-East Europe should have happened through the large conquests carried out in the Ancient Greek and Persian empires. Besides the mythology and religious aspects, the use of indigo by the Ancient Egyptians indicates that the relevant extraction knowledge appeared early in human history. Even if the starting point is not clearly known, sources of indigo are, however, identified: indigo has been

extracted from plants using diverse methods depending on the plant family and the region of the world where it was exploited. As mentioned above, more than 300 plant species are able to produce indigo. Depending on the plant source, the color varies from the palest sky blue to the deepest midnight blue. For reasons of clarity, only the three main species that contributed the most to the development and reputation of indigo around the world will be described.

3.1 Sources and Extraction of Indigo Around the World

3.1.1 Japan, Asia

3.1.1.1 History

In Japan, the leaves of *Polygonum tinctorium* Lour. (Polygonaceae) have been used as a traditional source for indigo production. This plant was also used in other areas of Asia, especially in mainland China, along with *Isatis indigotica* Fortune. The production of indigo was known in China since the Zhou period (1045–771 BCE) and indigo (*sukumo* in Japanese) was exported to Japan in the fifth or sixth century (*24*). In both countries, it was used to dye traditional costumes such as the kimono or simply everyday clothes. Its cultural importance was so encrusted in the Japanese society that the Japanese Ministry of Culture in 1978 protected indigo as a "National Treasure" and the dyers know-how was regarded as a "Living Treasure". The Asian indigo-bearing plants were also utilized for another application in societies of the Far East, as the basis of a traditional medicine like Chinese *Danggui Longhui Wan* (see Sect. 5.1.1).

3.1.1.2 Extraction Process

Extraction of indigo from *P. tinctorium* reveals centuries of knowledge, with the tradition kept jealously by Japanese dyers. It consists of harvesting and fermenting leaves to increase the indigo precursor quantities. The whole procedure extending over the entire year has been described by *R. Ricketts* (*24*), and will not be discussed in detail in this contribution. The chemistry involved in the extraction process is included in the section on Indian plants (Sect. 3.1.3).

3.1.2 Europe

3.1.2.1 History

In Europe, indigo was produced predominantly from *Isatis tinctoria* L. (Brassicaceae), which is better known under the name of "woad" (Fig. 5). This biennial plant has been exploited in Western Europe since the Neolithic period (*25*), as attested to by the discovery of seeds in France. The Roman Emperor *Caesar* described

Pl. 38. Pastel des teinturiers. Isatis tinctoria L.

Fig. 5 Botanical representation of *Isatis tinctoria*. (The picture is in the public domain. Source: http://www.commons.wikimedia.org/wiki/File:38_Isatis_tinctoria_L.jpg)

the use of blue pigments by the Britons in England, but it has not been documented definitively that this blue color came from *I. tinctoria*.

The first textiles dyed with woad appeared during the Iron Age (~700 BCE) (*25*) and the beginning of the dye industry started during the Middle Ages. The main production areas of activity were concentrated in England (Lincolnshire and Somerset), Germany (Thuringia), Italy (Piedmont and Tuscany), and France (Gascony, Normandy, Britany, the Somme Basin). However, the main European area dedicated

to indigo production was undoubtedly the Languedoc area in France, which led to some cities such as Toulouse, Albi, and Carcassone becoming extremely rich.

3.1.2.2 Extraction Procedure

Traditional extraction of indigo was performed from the plant leaves. After being harvested, the leaves were crushed and mixed with water to make a pulp and compressed into a pellet form (*cocagne*). After drying, the pellets were crushed into a powder and mixed with urine to induce oxidation. This oxidative stage leads to a paste, which once dried affords the dye or *pastel*, containing indigotin (**8**).

Currently, the extraction is conducted according to three major phases: maceration, oxidation, and precipitation. The leaves are immersed in a water bath (maceration phase) to extract indigo precursors; and the latter are then oxidized by stirring the bath (oxidation phase). The color of the water then switches from green to deep blue. The liquid is then allowed to stand and the dye is collected at the bottom of the bath by precipitation.

3.1.2.3 Chemistry

Indigo is not produced directly by woad. It is the result of a series of chemical reactions and several studies (*16*, *26–29*) have identified these indigo precursors. The major precursor of indigo is isatan B (**23**) (indoxyl-β-ketogluconate). The minor precursors, indican (**24**) (indoxyl-β-D-glycoside) and isatan C (**25**) (*16*) (Fig. 6), have also been identified.
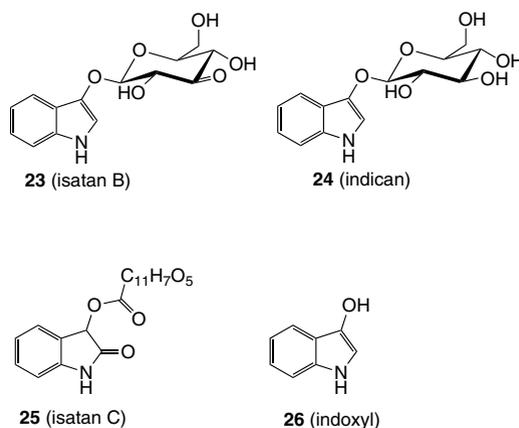
Fig. 6 The indigo precursors

During the traditional preparation, the precursors are degraded by hydrolases released when the leaves are crushed. The liberated indoxyl (26) is then oxidized by the addition of urine to form the *pastel*.

Since the leaves are not crushed during the modern extraction procedure, isatan B (24), indican (25) and isatan C (26) are solubilized in the water bath. The maceration creates a medium rich in bacteria and sufficiently alkaline to break the glycosidic bond and liberate indoxyl (26). Under the influence of both aerial oxygen and basic conditions, indoxyl dimerizes to form indigotin (8) during the stirring step. However, indigotin (8) is not the only compound to be formed during the extraction process. Considering the nature of the precursors, side reactions occur and lead to the formation of indirubin (9), isoindigo (11), and isoindirubin (12) (*16*), as described in Scheme 3.
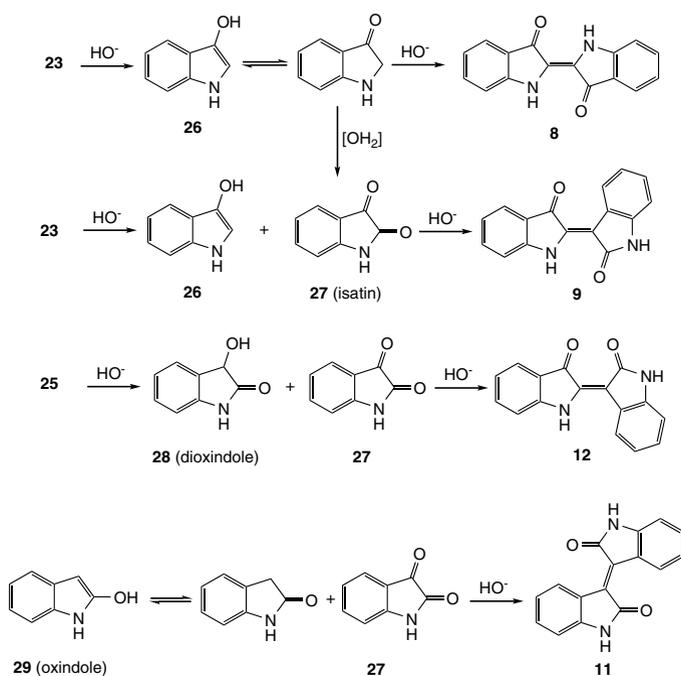
Scheme 3 Indigoid derivatives formed during the woad extraction process

Indigotin (8), the main coloring agent, is obtained by the dimerization of isatan B (23). Indirubin (9), the principal minor compound, is formed by the coupling of indoxyl (26) and isatin (27) with the latter resulting from the oxidation of indoxyl (26), due to the stirring conditions or on addition of urine. The decomposition of isatan C (25) leads to the formation of dioxindole (28), which can be oxidized to isatin (27). The dimerization of these two compounds affords isoindirubin (12). Finally, as a minor reaction, oxindole (29), an isomer of indoxyl (26), reacts with isatin (27) to yield isoindigo (11).

The biogenetic precursors of isatan B (**23**) and indican (**24**) have been studied by feeding young *Isatis tincotria* plants with [H³]-labeled-tryptophan and [C¹⁴]-labeled-acetate (*30*, *31*). Indican (**24**) and isatan B (**23**) labeled on their indole (H³) and glucosidic parts (C¹⁴) have been isolated and identified, thereby proving their origin from indole. The shikimic acid pathway is probably involved in their biosynthesis (Scheme 4).
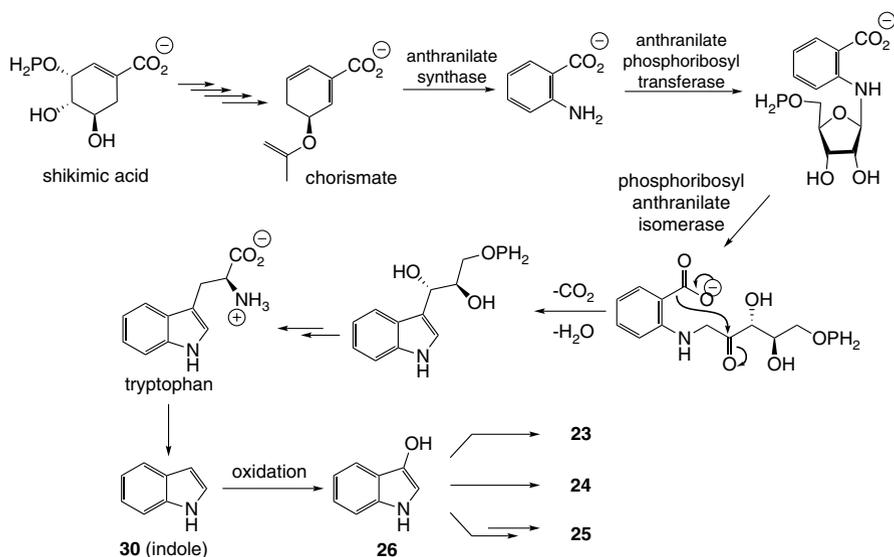
Scheme 4 Biosynthesis of indigo precursors

However, the pathway of the formation of indoxyl (**26**) from tryptophan is still obscure. Nonetheless, the biochemical synthesis of indigotin (**8**) from strains of *E. coli* (*21*) suggests that indigo precursors formation occurs through tryptophanase and dioxygenase.

The harvest time, the treatment, and the soil impurities (*32*, *33*) affect the proportion of the constituents and so the resultant color. This could explain the *pastel* color differences observed all over Europe and moreover the assured quality of the French indigo. However, at the dawn of the sixteenth century, woad production in France was curtailed due to the arrival of a darker indigo along with the discovery of a sea route to India by *Vasco de Gama*.

3.1.3   India, South America, and French West Indies

3.1.3.1   History

From the beginning of the sixteenth century to the seventeenth century, the European *pastel* was supplanted by the more intense indigo from India. It was extracted mainly from *Indigofera tinctoria* L. (Fabaceae) or "true indigo". This two-meter

high tree is well acclimated to the tropical areas from South Asia and can be found in some parts of Africa. However, its original habitat is not known and according to the climate, it could be an annual, biennial, or perennial. More than 800 species of *Indigofera* exist, but most of them are not of interest for the production of dye. *Indigofera tinctoria* L. (India, Southeast Asia, Middle East Africa, Madagascar) and *Indigofera suffruticosa* Mill. (South America, French West Indies (Fig. 7)) remain the most widely used plants in this regard.

Fig. 7 Botanical representation of *Indigofera suffruticosa*. (The picture is in the public domain. Source: http://www.commons.wikimedia.org/wiki/File:indigofera_suffruticosa_-_K%C3%B6hler%E2%80%93s_Medizinal-Pflanzen-076.jpg)

As one of the first suppliers of the Greco-Roman world, the dye industry in India had an important impact in society and economy. Except for exportations, blue-dyestuffs were manufactured for everyday life, or for ceremonial or religious

purposes. The production of Indian indigo exploded when the Dutch East India Company and the British Empire took the control of India and its raw materials. The development of shipping routes to South America, as effected by the Spanish and the Portuguese kingdoms, marked the starting point of the introduction of indigo from *I. suffruticosa* on the European market. The commercialization of Indian and American indigo dislodged the French woad and initiated an economic war between the colonial empires, mainly between France and Great Britain. Due to the sugar crisis in nineteenth century, France relocated the production of indigo to the French West Indies (Fig. 8), on the islands of Marie-Galante (*34*) and Guadeloupe (*35*), where *I. suffruticosa* was present (or introduced).
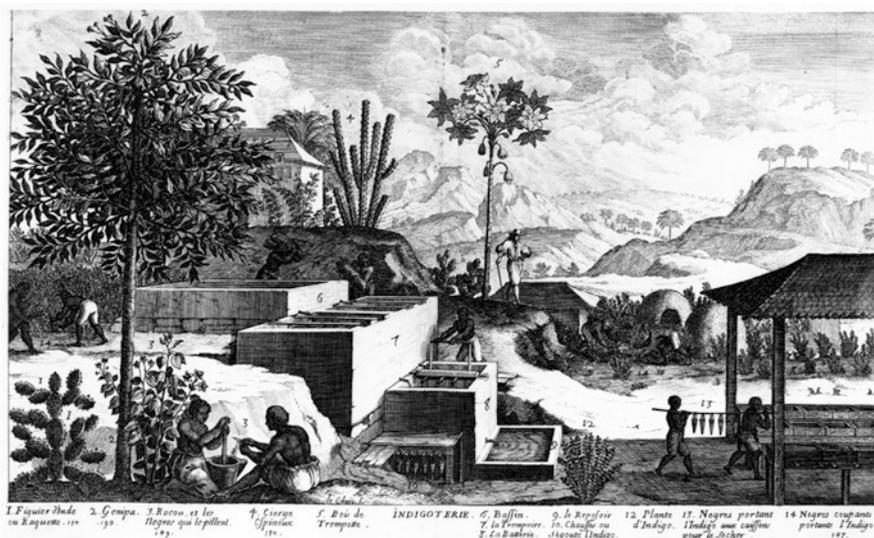
Fig. 8 Representation of indigo manufacture by *Dutertre* (seventeenth century) in the French West Indies. (The picture is in the public domain in US. Source: http://www.commons.wikimedia.org/wiki/File:indigoterie-1667.jpg?uselang=fr)

3.1.3.2 Extraction Process

Regardless of the place of production or the actual plant used, the extraction processes used for indigo samples have been quite similar. Thus, the leaves were placed in a large tank, maintained at the bottom by large wooden beams, soaked in water, and left to macerate. In India, slaves and later free workers entered the bath and stirred it with their legs or with wood sticks to create the correct oxidative conditions. In South America, workers manipulated wooden paddles outside the bath to stimulate the stirring process. In the French West Indies, a system of communicating baths was created: after maceration, the water went down to another tank and was stirred by wooden blades (named *batterie* in French) that were activated manually by slaves (*34*).

In all cases, after the stirring process, the dye was recovered at the surface of the water. Since the extraction step is the same as that for woad, the chemistry involved is then identical. What is the difference between European and Indian/American/Caribbean indigo? This will be answered in the next section of this chapter.

3.1.3.3 Chemistry

Studies (*29*, *36–38*) of the plants used to produce indigo in India, America, Thailand, and Japan (*Polygonum tinctorium*) have shown that the main precursor is indican (**24**). The numbers of reactions possible is limited and thus so are the number of potential derivatives. While three different precursors are present in *Isatis tinctoria* and lead to the formation of four colored molecules, only one is present in the *Indigofera* species used, and this leads mainly to the formation of indigotin (**8**) and its isomer, indirubin (**9**) (Scheme 5). The concentration of indigotin (**8**) is then higher, thus resulting a more intensely blue and pure dye.
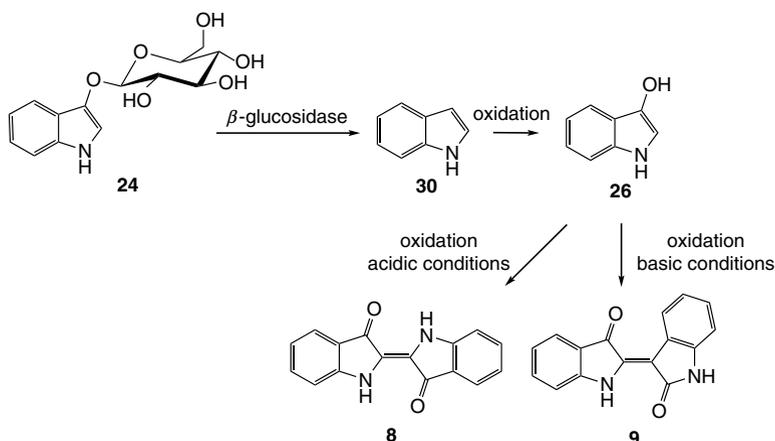
Scheme 5 Formation of the indigo and indirubin in *Indigofera* and *Polygonum* species

The struggle for the control of the indigo market lasted until the beginning of the nineteenth century. Independence movements in the different colonies, especially rebellious events occurring in India, diminished the external control by colonial empires. Moreover, the first chemical synthesis methods successfully adapted to an industrial scale ended trade in natural dyes and hence the prosperity of their local industries.

3.2   Chemical Synthesis of Indigo

Numerous syntheses of indigo are described in the literature. For reasons of clarity, this chapter will focus on the more relevant methods, starting from the first synthesis by *Baeyer*.

3.2.1 The First Syntheses: *Baeyer*, *Drewsen*, and BASF

The excessive cost in European markets of natural indigo dye, resulting from long transportation times, the emancipation of colonies, and the rise of the industrial revolution led to a reconsideration of the use of natural supplies. German chemists started developing chemical synthesis of indigo with the support of BASF and Hoechst, two important companies involved in the industrial development of indigo.

Adolf von Baeyer proposed the first synthesis of indigotin (**8**) from isatin (**27**) in 1870 (*39*) and its modification in 1879 (*40*) (Scheme 6). This method involved the chlorination of the 2-position of isatin by phosphorus chloride followed by dimerization under acidic-reductive conditions using hydrogen iodide (*39*) or Zn (*40*) in acetic acid (Scheme 6). However, isatin (**27**) was very expensive at this time, and *Baeyer* proposed in 1880 (*41*) a new pathway starting from the *o*-nitro-cinnamic acid (**31**) (Scheme 6). Although this method was patented, it was also too expensive to be exploited.
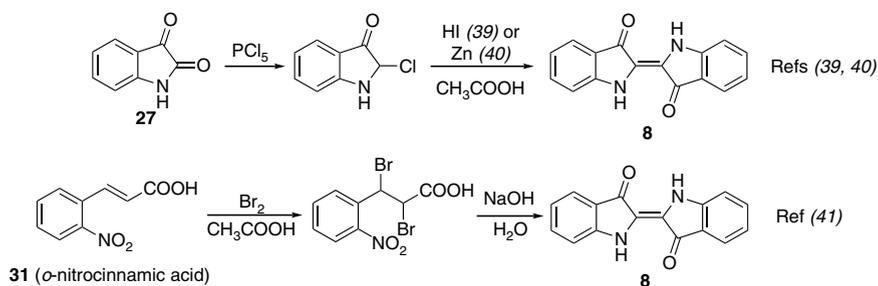
Scheme 6 First syntheses of indigotin (**8**) by *A. v. Baeyer*

A third method was described later by *Baeyer* and *Drewsen* involving *o*-nitro-benzaldehyde (**32**) (*42*). This synthesis pathway is known as the *Baeyer-Drewsen* method (Scheme 7).

However, this method was impractical because of the difficulty in making the starting product. A major breakthrough was made in 1890 when *Heumann*
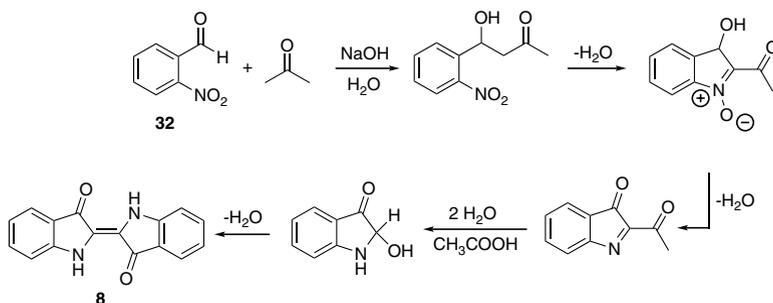
Scheme 7 The *Baeyer-Drewsen* method
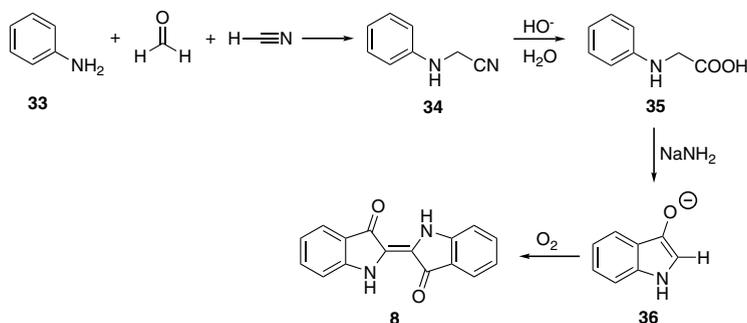
Scheme 8 Industrial synthesis of indigotin (**8**)

discovered the route to indigotin (**8**) from aniline (*43*). *Lucius* and *Brunning* then improved his method. To date, this synthesis (Scheme 8) dating from 1925 and introduced by BASF, is the most widely used method and is still utilized today. This pathway involves the condensation of formaldehyde and hydrogen cyanide to form the corresponding cyanhydrin. The latter is then attacked by aniline (**33**) to form the corresponding 2-(phenylamino)acetonitrile (**34**). The hydrolysis of the cyano group leads to the β-aminoacid (**35**) being fused into indoxylate (**36**) by sodium amide. The oxidation step gives indigotin (**8**).

3.2.2 New Methods and Mechanism Description

Diverse methods have been developed during the twentieth century (*44*). Most of these start from aniline (**33**), like the *Lucius-Brunning* method. One method of note involves the dimerization of indoxyl acetate, and was described in 1958 (*45*). It still has impact nowadays and an improvement has been applied in indirubin synthesis (Scheme 9).

The *Sandmeyer* method starting from aniline (**33**) leads to isatin (**27**) *via* the formation of the intermediate isonitrosoacetanilide (**37**). Isatin (**27**) is then oxidized in anthranilic acid (**38**). The latter is engaged in the *Lucius-Brunning* reaction to afford diacetylindoxyl (**40**) through the formation of phenylglycine-*o*-carboxylic acid (**39**). Deacylation in alkaline conditions followed by the indoxyl acylation leads to 3-acetoxyindole (**41**). The formation of indigotin (**8**) is finally achieved by the dimerization of **41** in oxidative and basic conditions.

A study reported by *Kaupp*'s group (*46*) showed that the coupling reaction of indoxyl acetate in an alkaline medium generates an indoxyl radical (**42**) due to oxidative conditions. Two indoxyl radicals coupled to form leuco-indigo (**43**), a colorless compound. Leuco-indigo (**43**) is then quickly oxidized by air in indigotin (**8**) (Scheme 10).

The chemical synthesis of indigotin (**8**) led to the rapid development and internationalization of the textiles industry. The blue jeans production introduced earlier in
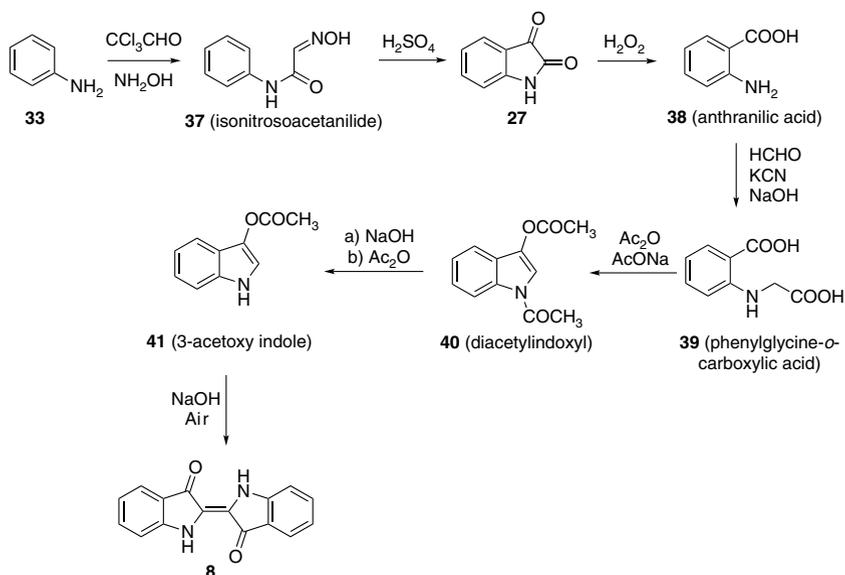
Scheme 9 Method based on the dimerization of 3-acetoxyindole (**41**)
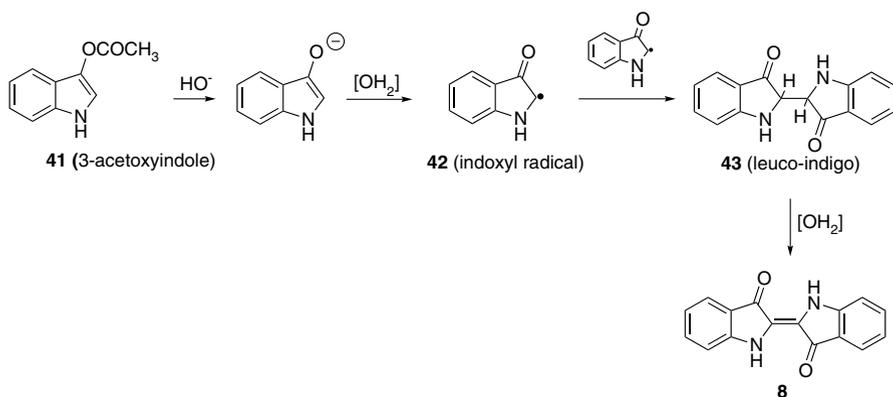
Scheme 10 Mechanism of indigo formation

America by Levi Strauss with denim indigo (produced in Nimes, France) has shown a great renewal. Nowadays, the global annual production is estimated at a billion pairs of jeans. The dyeing process involves the reduction of indigo to form the water-soluble leuco-indigo (**43**). Then, the cloth is steeped in the dye-containing water. On removal from the bath, subsequent oxidation in air colors the fabric blue.

Indigo has had great influence on human society. Its influence was so important that it led empires to fight to take over its control and to generate economic benefits.

As the industrial revolution took place, indigo led to its own revolution through the development of the chemical synthesis methods and the rapid application by the textiles industry. Nowadays, new indigotin (**8**) derivatives have been synthesized and have found applications in medicinal chemistry.

4 Tyrian Purple

4.1 Legendary History

As I have no desire to tire you with endless lessons on a single subject, I will tell you instead the origins of purple […]. Hercules had a dog with him, as was the custom in the ancient world; as you know, dogs accompanied the heroes of old even when they entered an assembly. Hercules' dog saw a purple mollusk crawling up a rock, poking out of its shell; he seized his flesh in his teeth, then ate it. Its blood dyed the dog's lips the brightest red … The nymph, when Hercules was on her side, saw the dog with his lips this unusual color, and declared that she would refuse Hercules her love unless he presented her with garments even brighter than the lips of his dog. Hercules went back for the shell, extracted the blood, and gave the girl the gift she so desired, thereby acquiring the reputation in Tyre of inventor of the purple dye (Julius Pollux, 2nd century BCE) (47) (Fig. 9).

Tyrian purple (Royal Purple, Imperial Purple, or Purple of the Ancients) is a natural purple dye for which the extraction and distribution was of economic and

Fig. 9 Discovery of Tyrian purple by *Hercules* (Ref. (*48*))

social importance in the Mediterranean Sea, starting from Antiquity to the fifteenth century of the Common Era. Although legend locates the first extraction in Tyre (Lebanon), documented evidence from the Minoan period (twenty-seventh–fifteenth century BCE) could place its first extraction in Crete (Greece). Undoubtedly, its first use took place during the seventeenth century BCE in Akrotiri of Thera (Santorini Island, Greece) (*49*, *50*), as proved by HPLC-APCI analysis (*51*). Wall paintings discovered in this prehistoric city testified to its wealth during the Minoan Period. There is an evidence for the Royal Purple industry around the thirteenth century BCE in Serafand (Lebanon) (*52*). Its extensive production from the Phoenicians in the Tyre area dates back to 560 BCE and Tyrian purple has been widely used in Persia by the Achaeminian dynasty (559–330 BCE) (*9*, *53*, *54*). The meteoric rise of the dye justified its price: it was worth 10–20 times as much as its weight in gold (*52*). The blue Tekhelet, a variation of Tyrian purple mentioned in the Jewish bible, was used to dye the cord of the High Priest (*52*). After the conquests achieved by *Alexander the Great* (331 BCE) the eye-opening dye became well known in the north of Greece, especially in Macedonia (*9*). The beautiful color and its expensive value pushed the Roman emperor *Nero* to impose sumptuary laws and restrict its use for the ruling classes only, thereby becoming the symbol of power. Later, this law was extended to ceremonial purposes and the robes of popes and cardinals were also dyed with Tyrian purple. The fall of the Roman Empire slowed its production. Later, the fall of Constantinople in favor of the Ottoman Empire (1453 CE) and the decision of Pope *Paul* II to replace purple by carmine (1464 CE) (*55*) brought a fatal blow to the production of purple in Europe. *William Cole* from Bristol later rediscovered Purple of the Ancients in 1684 (*56*) on the shores of England. Nowadays, purple is still traditionally extracted in South America and its renewal has emerged in arts in Europe. 6-Bromoindirubin (**13**), one of the constituents of the dye, is actually a lead compound in medicinal chemistry research, thereby prolonging the legendary history of Tyrian purple.

4.2 Extraction

4.2.1 Sources

The purple dye is extracted from number of mollusks belonging to the family Muricidae. This worldwide family groups a wide variety of shaped and colored shellfish. Dye production is a characteristic of this conchylian family. In southeastern Europe, the highly prized Tyrian purple was obtained at great expense essentially from the two species *Bolinus* (*Murex*) *brandaris* and *Hexaplex trunculus*. It has been reported that 10,000 animals were needed to produce 1 g of the dye (*54*). In Britain and Ireland, the rediscovery of purple by *W. Cole* was possible owing to the occurrence of *Nucella lapillus* (dog whelk). *Thais kiosquiformis* (South America), *Plicopurpura* species (Mexico, Caribbean Sea, Atlantic Ocean), and *Rapana venosa* (Japan) have been subjected to dye production in their respective areas (*55*).

4.2.2 Extraction Process

The first surviving details of a purple dye existence were found on a Mycenaean Linear B tablet that dated from the thirteenth century BCE (*50*). *Pliny* the Elder was one of the first naturalists to describe the extraction in *Naturalis Historia* (*57*). The dye does not exist in the mollusk as indigo, as in an indigo-bearing plant. The purple color is generated from precursors, the so-called "chromogens", contained in the hypobranchial gland. To access the vein, the shell should be broken carefully so as not to crush the organism inside. Once the vein is accessible, a colorless liquid can then be extracted, which once exposed to light becomes purple in color.

4.2.2.1 Central America-South America-Caribbean Sea

Thais kiosquiformis and *Plicopurpura patula pansa* were mainly used. Traditionally, the shellfish was impaled with a cactus needle and pressed to collect a milky liquid on a cotton piece. The shellfish was then allowed to stand on a rock while waiting to be extracted a second time. The organism was then withdrawn from the shell and compressed with a knife from head to tail and the liquid collected on a piece of cotton. This process is somewhat reminiscent of that used by farmers to take milk from cows: it has thus been baptized "shellfish milking". Sometimes this operation was repeated three or four times before the death of the animal.

4.2.2.2 Japan

Few details have appeared in literature concerning purple dye extraction in Japan. It is known that the dye was extracted from *Rapata venosa*. The numbers of broken shells found in the area of Hitachi (*55*) attests to a similar process used as described in Sect. 4.2.2.1. Even though published information is scarce, traces of the dye have been identified on silk from Yoshinogari (Kyushu Island) (*58*).

4.2.2.3 Europe

H. trunculus and *B. brandaris* are the two most abundant species used in countries bordering the Mediterranean Sea in Europe and North-west Africa. *N. lapillus* is found preferentially in northern Europe (Atlantic Ocean). These three shellfish have been used widely used and two distinct methods have been described for dye production. The first one described by *Pliny* the Elder is an indirect extraction/dye process:

> *"The vein already mentioned is then extracted and about a sextarius of salt (3.17 kg) added to each hundred pounds (0.453 kg) of material. It should be soaked for three days, for the*

fresher extract, the more powerful the dye, then boiled in a leaden vessel. Next, five hundred pounds of dye-stuff, diluted with an amphora (30.28 L) of water, are subjected to an even more moderate heat by placing the vessels in a flue communicating with a distant furnace […] and a test is made about the tenth day by steeping a well-washed fleece […] The Tyrian color is obtained by first steeping the wool in a raw and unheated vat of pelagian extract, and then transferring to one of buccine…" (*Pliny* the Elder, *Historia Naturalis*, 1st century CE) (*59*).

The series of heating and dilution steps in differently coated containers implies that a series of chemical reactions was used to obtain the best dye (see Sect. 4.2.3.4). The processing of *N. lapillus* for dyeing was similar to that used in South America. The shell was broken and the vein penetrated. The viscous liquid was applied directly onto linen and left in the sun (*56*).

4.2.3 Chemistry

4.2.3.1 Composition

The main component of interest is 6,6′-dibromindigotin (**17**), and in some species such as *R. venosa* or those in the genus *Plicopurpura*, it can be found alone (*58*, *60*, *61*). It is often called "Tyrian purple" by reference to the dye and represents the real biomarker of true-purple dyes. The second main component is 6,6′-dibromoindirubin (**15**), which has been identified in *H. trunculus* (*9*, *62*), *N. lapillus* (*63*), and *B. brandaris* (*64*). Two other brominated indirubins have been discovered: 6-bromoindirubin (**13**) and 6′-bromoindirubin (**14**). Traces of 6-bromoindigotin (**16**), indirubin (**9**), and indigotin (**8**) also have been found in *H. trunculus* (*61*, *62*).

4.2.3.2 Precursors and Biosynthesis

Conchylian indigoids are formed next in a series of chemical reactions from precursors present in the hypobranchial glands. Identification of these precursors (Fig. 10) was a high scientific objective for dye chemists. It is suggested that four chromogens are always present in species in the family Muricidae, thanks to the contributions of *Baker* (*65*, *66*) and *Fouquet* (*67*). These are indoxyl sulfates (**44** and **45**) and brominated indoxyl sulfates (**46** and **47**). Baker has found tyrindoxyl (**48**) in the glands of *Dicathais orbita* (*65*). However, paper chromatography (*68*) has shown that a wide variety of indoxyl sulfates exist in shellfish species. *Baker* and *Duke* have isolated a series of choline (**50**) and choline ester salts of tyrindoxyl sulfate (general structure (**49**)) from *D. orbita* and *Mancinella keineri* (*69*). Murexine salt (**51**) has been found to be the ultimate precursor of Tyrian purple in these species. Nonetheless, precursor differences have been observed according to the season (*61*) and the sex gender of the mollusk (*70*) making the identification more difficult.

The biosynthesis pathway leading to the formation of the precursors is not fully understood, but is assumed to be similar to that of indigo. The precursors may be
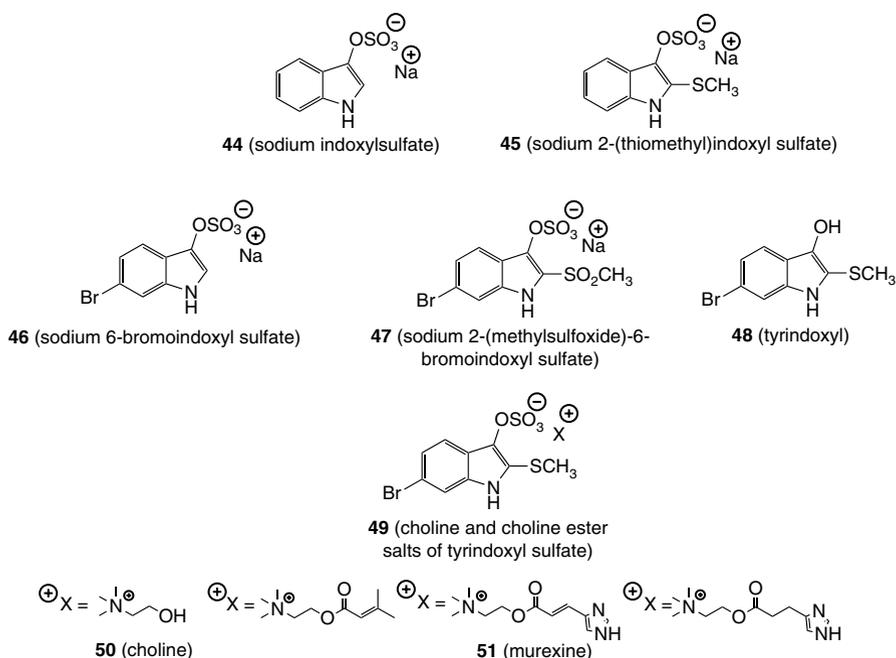
Fig. 10 Identified Tyrian purple precursors

formed from tryptophan through the action of a tryptophanase and/or other enzymes to indole (**30**). Indole (**30**) can then be oxidized by a diooxygenase to indoxyl (**26**), with the latter being converted to the corresponding sulfate. Alternatively, the pathway may convert tryptophan directly to indoxyl sulfate (**44**), following a new biosynthesis pathway. However, the origin of tryptophan or indoles in the mollusks could also be external to these organisms. Muricidae shellfish possess the ability to biosynthesize choline and its ester derivatives. Up to the present, no experimental proof has been obtained concerning their biosynthetic origin. The specific positions of bromine and sulfate groups suggest the need for specific enzymes. The absence of bromotryptophan in the hypobranchial glands (*71*) and evidence for post-translational bromination at the 6-position of tryptophan in a sea snail (*Conus radiatus*) (*72*) support this hypothesis. Moreover, bromoperoxidase and vanadium bromoperoxidase have been found in some algal species (*73*) and could be the enzymes involved (Scheme 11).

Although the biosynthesis remains elusive, the precursor function of the mollusk has been clarified thanks to investigations by *K. Benkendorff*. She proposed a role in reproduction and defense for the indole derivatives and a *de novo* synthesis of indigoids inside the mollusk (*71*, *74*).
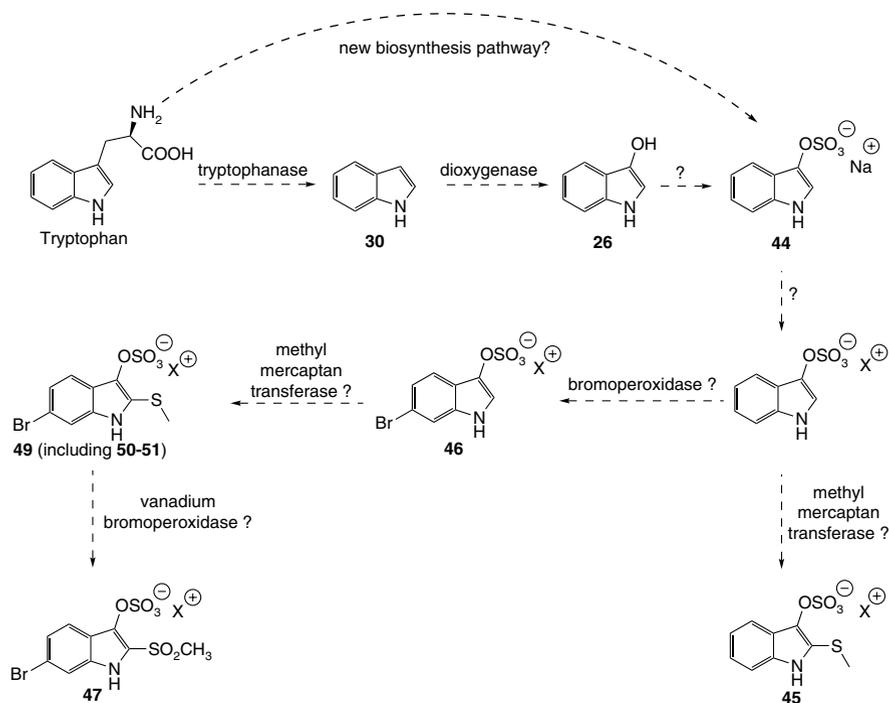
Scheme 11 Hypothetic biosynthesis of Tyrian purple precursors

4.2.3.3   Chemical Process

When exposed to light, these precursors undergo chemical and enzymatic transformation easily observed by applying the dye on fibers (Fig. 11). Several investigators have described color development from a white liquid becoming successively yellow, green, and then purple.

The nature of the intermediates has been the object of scientific debate (*65*, *75*). However the formation of 6,6′-dibromoindigotin (**17**) and 6,6′-dibromoindirubin (**15**) is now accepted (Scheme 12).

Murexine-tyrindoxyl sulfate (**51**) (*70*) (white) undergoes a desulfonation to form tyrindoxyl (**48**), by action of a purpurase identified as arylsulfatase (*67*). Tyrindoxyl (**48**) under oxidative conditions leads to tyrindoleninone (**52**). Coupling between **48** and **52** affords the green intermediate tyriverdine (**53**) (*76*). Photolysis of the two methanethiol moieties yields 6,6′-dibromoindigo (**17**). 6,6′-Dibromoindirubin (**15**) is formed by the coupling reaction of tyroxindole (**48**) and 6-bromoisatin (**54**). The latter is generated by both oxidation of tyrindoleninone (**52**) and tyriverdine (**53**).
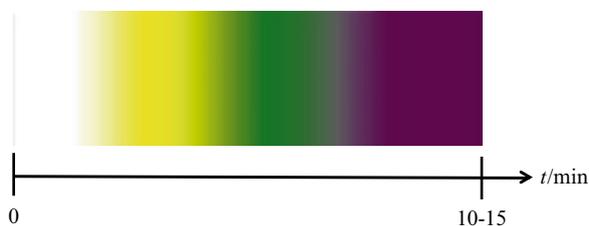
Fig. 11 Development of the color
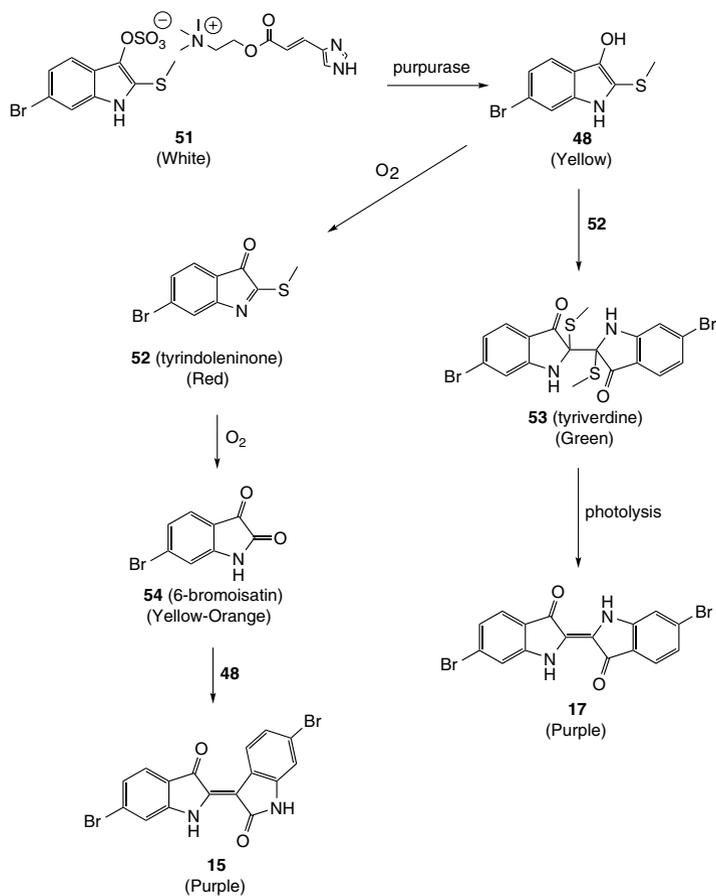
Scheme 12 Formation of 6,6-dibromoindigotin (**17**) and 6,6′-dibromoindirubin (**15**)
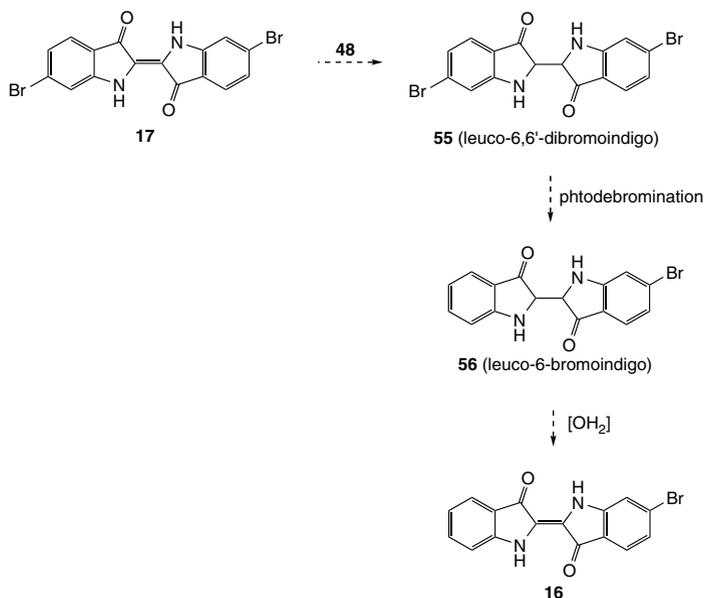
Scheme 13 Formation of 6-dibromoindigo (**16**)

Formation of monobrominated indirubins and 6-bromoindigo (**16**) is, however, still obscure and there is no clear evidence of their formation to date. 6-Bromoindigo (**16**) probably arises from photodebromination of leuco-6,6′-dibromoindigo (**55**) formed by reduction of 6,6′-dibromoindigo (**17**) by tyrindoxyl (*68*). Leuco-6-bromoindigo (**56**) is then oxidized to give 6-bromoindigo (**16**) (Scheme 13).

The same mechanistic explanation is given for the formation of indirubin (**9**) (Scheme 14). 6,6′-Dibromoindirubin (**15**) is reduced by tyrindoxyl (**48**) to form leuco-6,6′-dibromoindirubin (**57**). The latter undergoes a photodebromination to afford two leuco- intermediates possessing a bromine atom at position 6 or 6′ (**58** and **59**). These intermediates are separately oxidized to give 6′-bromoindirubin (**14**) and 6-dibromoindirubin (**13**). Nonetheless, this hypothesis has not yet been proven experimentally.

Another hypothesis highlights the presence of non-brominated precursors in *H. trunculus* (Scheme 15). Sodium indoxyl sulfate (**44**) may be transformed in indoxyl (**26**) by action of purpurase. Indoxyl (**26**), by action of oxygen, could be transformed to isatin (**27**). Coupling between indoxyl (**26**) and 6-bromoisatin (**54**) affords 6-bromoindirubin (**13**) while the coupling of isatin (**27**) and tyrindoxyl (**48**) forms 6′-dibromoindirubin (**14**). As a minor reaction, indoxyl (**26**) and isatin (**27**) form indirubin (**9**) and the dimerization of indoxyl (**26**) leads to indigotin (**8**).
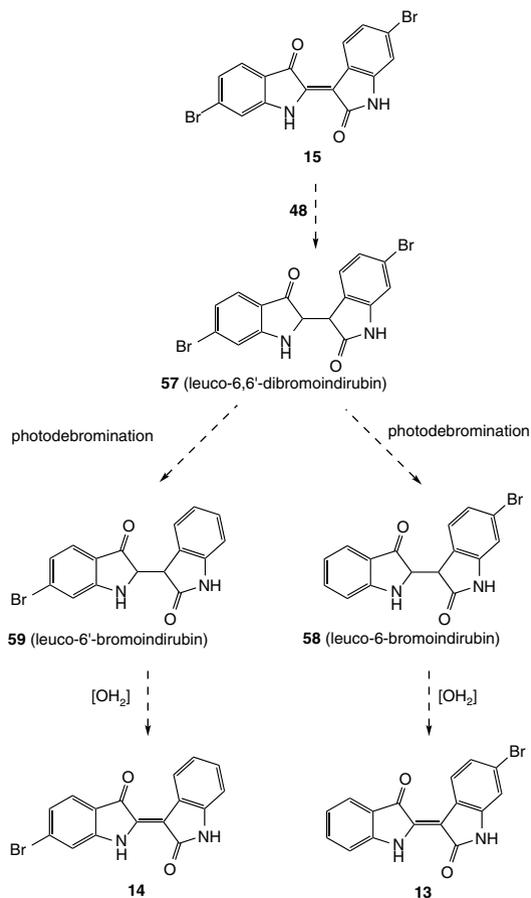
Scheme 14 Formation of mono-brominated-indirubins
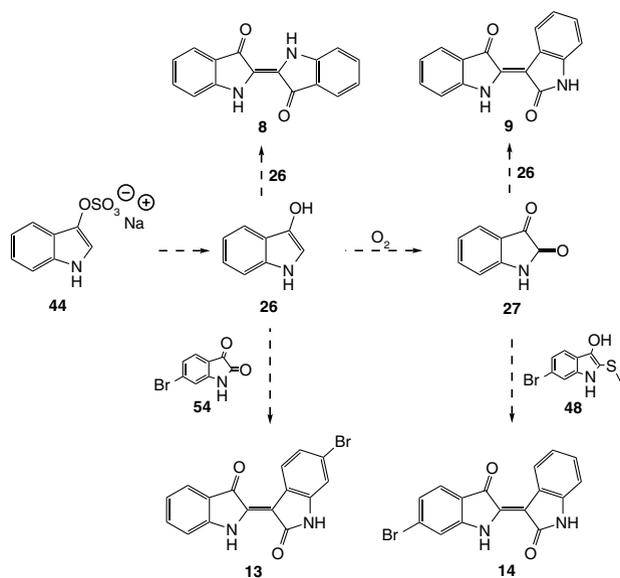
Scheme 15 Second hypothesis on formation of mono-brominated-indirubins

4.2.3.4 The Case of Vat Dyeing

The Phoenician dye technique is reminiscent of vat dyeing, which has been described by *Pliny* the Elder (see Sect. 4.2.2). The dye was maintained in the leuco-form because of the water insolubility of 6,6′-dibromoindigo. The cloth was dipped inside and aerial oxidation led to the purple hue. A series of chemical reactions is involved and these are linked with dilution steps and differential heating in a specially coated vat. This implies the establishment of a reductive system, precursor stabilization, enzyme deactivation, and photochemical control. *McGovern* and *Michel* published a series of studies investigating the chemistry involved (*54*, *59*, *77*) that are summarized below.

> *"The best time to catch them is after the rising of the Dog star [...] The vein already mentioned is then extracted and about a sextarius of salt (3.17 kg) added to each hundred pounds (0.453 kg) of material. It should be soaked for three days, for the fresher extract, the more powerful the dye, then boiled in a leaden vessel."*

Pliny described the mollusks being collected during a specified period resulting therefore in the presence of specific precursors. The boiling step must surely deactivate the purpurase (deactivated at 75 °C (*59*)). However, in Latin, *Pliny* used the word *plumbo* translated into 'lead', but this could be understood as "tin" as well. *McGovern* and *Michel* thus performed reduction in aqueous solution of 6,6′-dibromoindigo (**17**) (DBI) in the presence of different metals to form the corresponding leuco-DBI (**56**) (*77*). To control the efficiency of reduction, they sank a piece of filter paper into the liquid and observed the resultant color. They discovered that tin and lead can both reduce DBI but only *tin in alkaline conditions (NaOH 1 N) at 90 °C* could give the deep purple characteristic of Tyrian purple. However, *Pliny* never mentioned the use of alkalis despite the fact they were known at that time. Moreover, the additives described (salt, urine) are not alkaline enough to reduce DBI (**17**) under these conditions (*59*).

In addition, the reductive agent might be present in the mollusk (as 2-substituted indoxyls (*59*)) or formed during the process (*e.g.* methylmercaptan from tyriverdine photolysis (*77*)). Experience using 1-dodecanethiol at about 80 °C has partially confirmed this hypothesis (*77*). Interestingly, *McGovern* and *Michel* noted that a natural product present in *B. brandaris* collected in France was highly efficient in reducing indigoids without any alkaline conditions (*77*). Unfortunately, this product has not been identified yet. However, the concentrations of methylmercaptan and 2-substituted indoxyls are low and the addition of honey (attested to by ancient texts and notably by *Plutarch*) could preserve the reduced-form of the dye as proved by a reduction trial using dextrose.

> *"Five hundred pounds of dye-stuff [...] are subjected to an even more moderate heat by placing the vessels in a flue communicating with a distant furnace [...] a test is made around the tenth day"*.

Stabilization of indoxyls must also be preserved and their oxidation to the insoluble indigoids blocked. At 40–50 °C and for long periods, the conditions described by *Pliny*, only the use of pure tin could prevent precipitation of the dye, confirming the use of tin-coated vat.

The authors showed that dark or subdued light could block the photolysis of diindoxyl. Several hypotheses could be proposed: the vat was covered, the dyers worked during the night or the insoluble by-products (other indigoids) formed a protective layer at the surface of water. Surprisingly, the text of *Pliny* does not mention the control of light and only one text refers to the effect of light without providing more precise details (*59*).

> *"The Tyrian color is obtained by first steeping the wool in a raw and unheated vat of pelagian extract, and then transferring to one of buccine [...] Conchyliated garments are prepared by a similar method to the first but [...] the dye is diluted simultaneously with urine and water".*

Here, "pelagian extract" corresponds to the *Murex* shellfish and "buccine" corresponds to an extract of *Buccinum* species (mollusk), which gives a red color. The famous Tyrian color was in fact a mix of two dyes, with the pelagian one serving to fix the buccine one.

All these studies were performed in the 1990s and do not answer all the questions. No recent studies have emerged since 1990. Additionally, difficulties are evident in chemically translate ancient texts. The decryption of the vat dyeing process of Tyrian purple used by the Phoenicians still remains to be chemically improved. However, the actual knowledge can be summarized as depicted in Scheme 16.
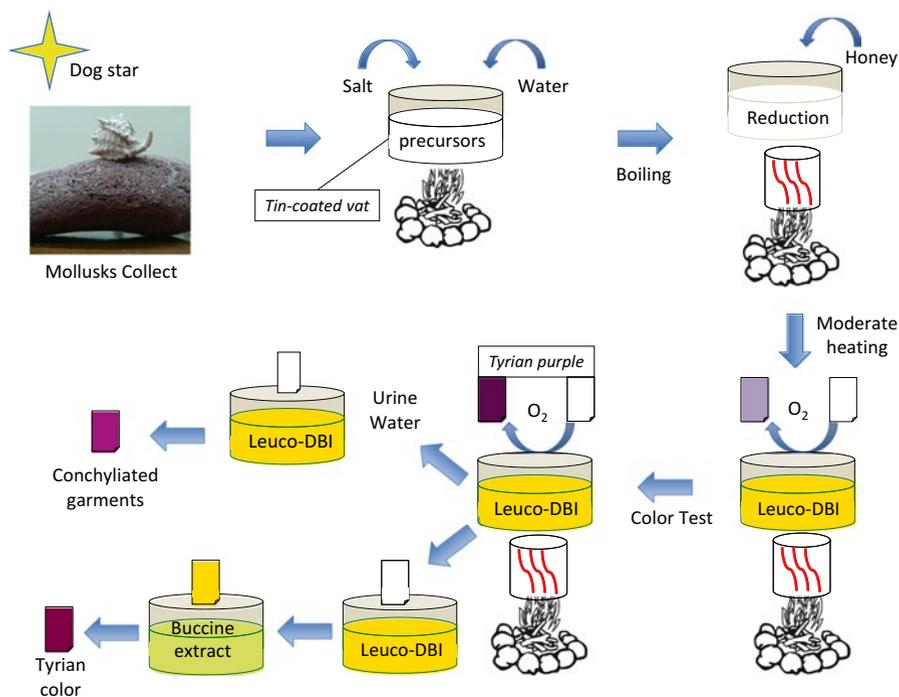
Scheme 16 Vat dyeing process

4.3   Chemical Synthesis of Tyrian Purple

The chemical synthesis of 6,6′-dibromoindigo (**17**) has been as great a challenge as the synthesis of indigotin (**8**). While its structure was not totally elucidated, chemists first explored its synthesis to aim for the formation of new indigotin derivatives. *Sachs* and *Kempf* were the first chemists to propose a total synthesis (*78*) (Scheme 17).
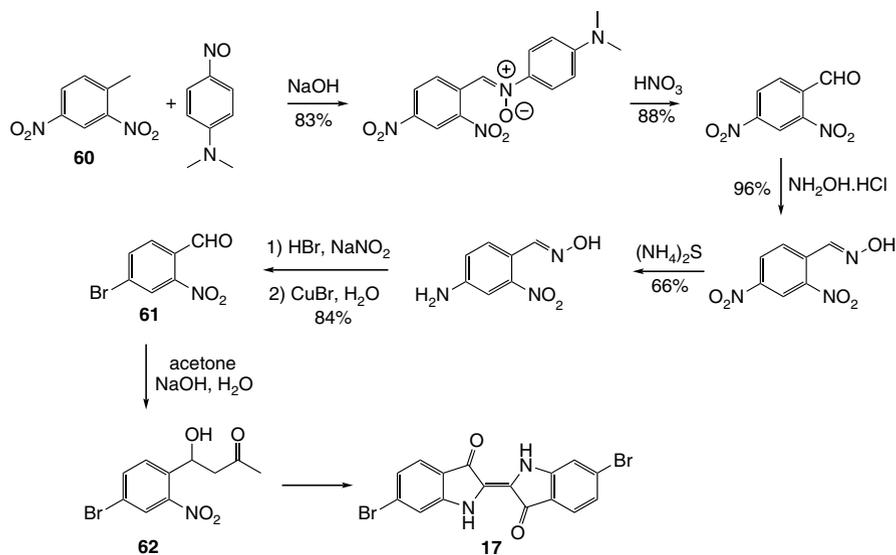
Scheme 17 First proposed synthesis of Tyrian purple by *Sachs* and *Kempf*

The strategy involves the transformation of 2,4-dinitrotoluene (**60**) to the key intermediate 3-bromo-6-formyl-nitrobenzene (**61**) in five steps. The bromonitrobenzene is then submitted to an aldolization in the presence of acetone in alkaline medium. The β-hydroxyketone (**62**) encounters ring closure to form an indoxyl derivative. The latter dimerizes spontaneously to afford 6,6′-dibromoindigo (**17**). The major inconvenience of this synthesis is the number of steps to form the key intermediate. Moreover, the authors did not mention the yield for the last step.

Later, *Sachs* and *Sichel* (*79*) proposed a slight modification to improve the yield of the dimerization. In this procedure, the key intermediate is condensed with acetone in the presence of trisodium phosphate and the β-hydroxyketone cyclized in alkaline medium. *Friedländer* was the first investigator to discover the structure of 6,6′-dibromoindigo (**17**) in 1909 (*80*). He reported that 12,000 shellfish (*M. brandaris*) were needed to provide 1.4 g of the dye. He proposed more convenient synthesis methods (*80*, *81*) starting from bromoanthranilic acid (**63**) (Scheme 18), with the latter being obtained from the oxidation of 2-amino-4-bromotoluene (**64**). Bromoanthranilic acid (**63**) reacts with chloroacetic acid to afford the corresponding phenylglycine (**65**). The following acetylation gives the *N*-acetyl-3-acetoxy-6-bromoindole (**66**). Alkaline hydrolysis under oxidative conditions leads to 6,6′-dibromoindigo (**17**). Although this method was more convenient than the *Sachs*
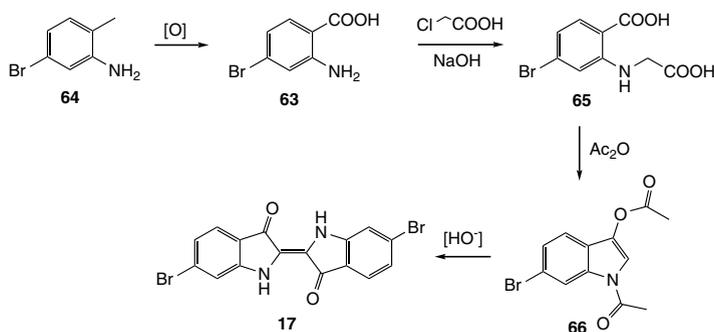
Scheme 18 *Friedländer* synthesis

strategy, the use of the expensive bromoanthranilic acid (**63**) as starting material represented an obstacle for industrial development.

As a consequence, *Grandmongin* and *Seyder* (*82*) proposed an alternative approach. 6,6′-Dibromoindigo (**17**) is formed by reduction of 6,6′-dinitroindigo (**67**), followed by bromination (Scheme 19).
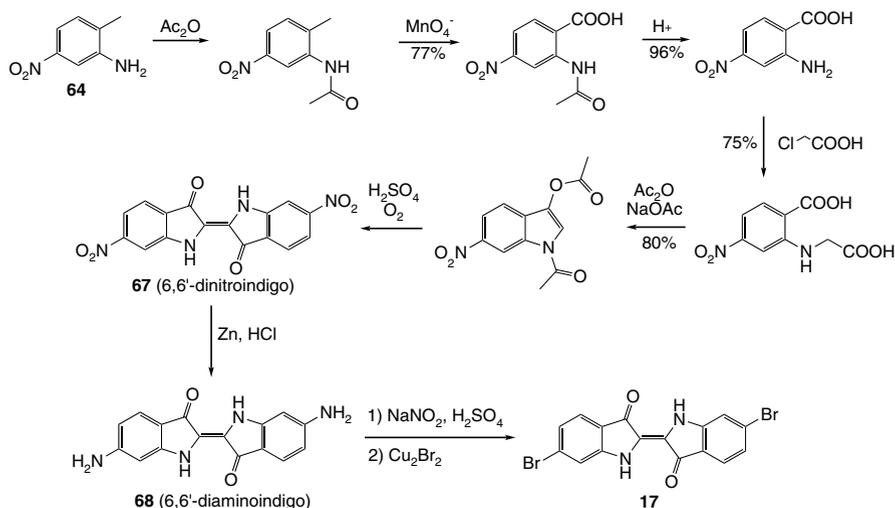
Scheme 19 Synthesis of 6,6′-dibromoindigo (**17**) according to *Grandmongin*

The 6,6′-dinitroindigo (**67**) is prepared in the same conditions as described by *Friedländer*. The two nitro groups are then reduced in the presence of zinc in acidic conditions and the 6,6′-diaminoindigo (**68**) thus formed reacts with sodium nitrite in acidic conditions to form the corresponding *bis*-diazo compound (not shown in the Scheme). The latter is then brominated by cupric bromide. This method is the only example of bromine introduction at the end of the synthesis.

With the exception of the *Grandmongin* method, all the procedures described share the common starting product 2-amino-4-bromotoluene (**64**), which is then
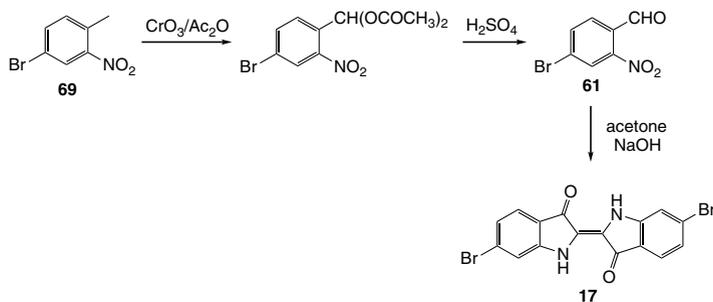
Scheme 20 *Torimoto* pathway
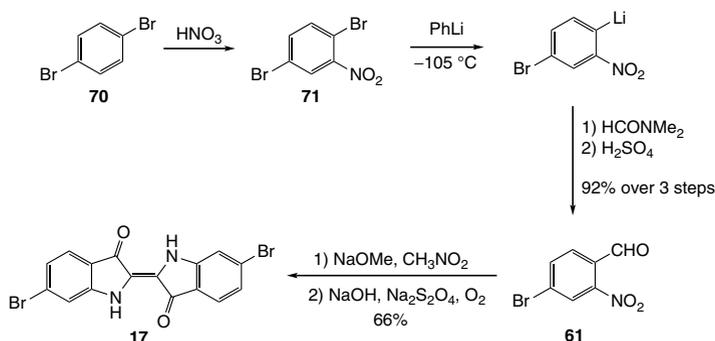
Scheme 21 Lithiation method developed by *Voss* and *Gerlach*

oxidized to the corresponding 3-bromo-6-formylnitrobenzene (**61**) or anthranilic acid. Therefore, synthesis of Tyrian purple has progressed through the perfection of the synthesis of intermediates. *Torimoto* cited in (*74*) proposed a new synthesis pathway (Scheme 20) to form 3-bromo-6-formyl-nitrobenzene (**61**) starting from 2-nitro-4-bromotoluene (**69**), thanks to the work of *Barber* (*83*).

Alternatively, *Voss* and *Gerlach* (*84*) proposed an approach starting from 1,4-dibromobenzene (**70**) based on selective lithiation at position-2 of 2,5-dibromonitrobenzene (**71**) to selectively introduce the formyl group in *ortho*-position of the nitro group to afford the key intermediate **61** (Scheme 21). The latter under basic and oxidative conditions affords the 6,6′-dibromoindigo (**17**). However, one of the major drawbacks of this method is to ensure a good control of temperature (−105 °C) for the lithiation.

Cooksey (*57*) developed a method to form the key intermediate by ameliorating the *Sachs-Kempf* method and applying it directly to the 2-nitro-4-bromotoluene (**69**) (Scheme 22). *o*-Nitro-toluene (**72**) is brominated at the *meta*-position of the nitro group by action of bromine in the presence of iron to form **69**. The methyl group is then halogenated by action of *N*-bromosuccinimide and the bromine further displaced by pyridine to form the corresponding pyridinium salt (**73**). The latter is converted to the nitroso derivative **74**, which by hydrolysis under strong acidic
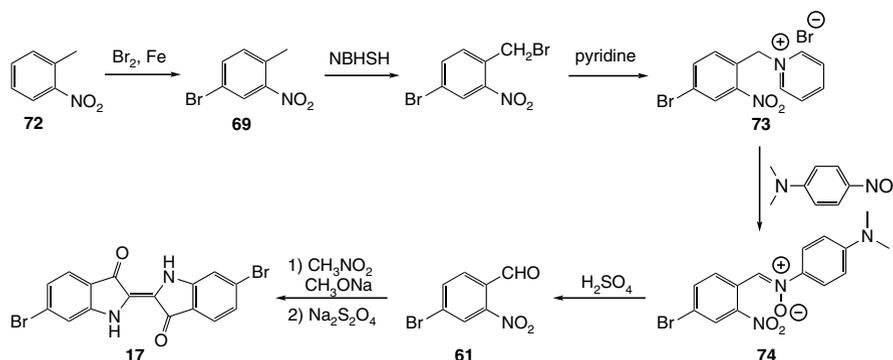
Scheme 22 *Cooksey* methodology
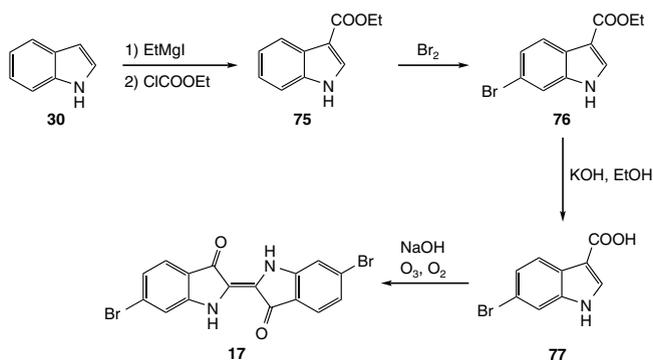
Scheme 23 *Majima* synthesis from indole (**30**)

conditions affords the corresponding aldehyde (**61**). The key intermediate is then dimerized in 6,6′-dibromoindigo (**17**).

Surprisingly, few methods involving the transformation of indoles have been developed, probably because of the cost of the starting materials at the beginning of the twentieth century. Later, *Majima* (*85*) and his collaborators have proposed a method starting from plain indole (Scheme 23). The successive treatment of indole (**30**) with a *Grignard* reagent followed by ethylchloroformate affords the indole-3-carboxyethyl derivative (**75**). Position 6 is then brominated and the ester (**76**) is converted to the corresponding carboxylic acid (**77**). Dimerization occurs under alkaline and oxidative conditions to furnish 6,6′-dibromoindigo (**17**). Although this method is attractive it suffers from the disadvantage of a lack of regioselectivity for the carboxylation and bromination steps (*86*).

Finally, *Tanoue* (*87*) has recently developed a method based on a biosynthesis pathway (Scheme 24). The 6-bromoindole (**78**) is converted in the corresponding 3-acetoxy-6-bromoindole (**79**) in a two-step method developed by *R. D. Arnold* (*88*). Position 3 is first iodinated and then acetoxylated by action of silver acetate in acetic acid. The 3-acetoxy-indole (**79**) is then dimerized in an alcoholic alkaline medium. This procedure is the fastest synthesis pathway but it is also the most expensive.
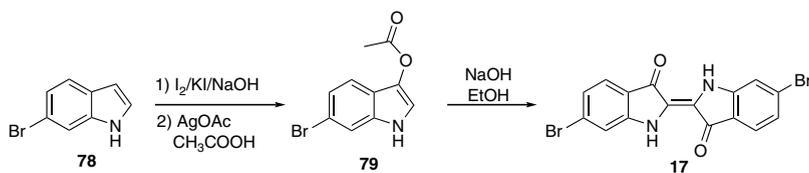
Scheme 24 Methodology of *Tanoue*

Tyrian Purple, as for indigo, has been the subject of numerous studies due to its perfect color and its legendary history, which together have pushed chemists to discover the complex origin of the color and to understand the sophisticated vat dyeing technique of the Phoenicians. The challenge to develop a sustainable synthesis for industrial applications was a major objective after the synthesis of indigotin (**8**). However, with the industrial development of western countries, the use of natural dyes has declined. Blue and purple became common colors, having social significance or not (blue has been the color of working class), and the chemistry of indigotin (**8**) and 6,6′-dibromoindigo (**17**) has been largely abandoned (except by dye chemists).

5   Renewal of Interest of Indigoids in Medicinal Chemistry

Despite the fact that the use of natural dyes declined over a period of time in favor of synthetic dyes, the indigoids, and especially the indirubins have been subjected to a renewed interest over the last ten years. Indeed, at the end of the 1990s, indirubin (**9**) was found to be the main active constituent of a traditional Chinese medicine called *Danggui Longhui Wan*.

5.1   Discovery of the Biological Potency of Indirubin

5.1.1   Danggui Longhui Wan

Danggui Longhui Wan has been used for the treatment of chronic myelogenous leukemia (CML) in mainland China (*62*, *89*). This medicinal formula consists of a mixture of eleven ingredients from traditional Chinese medicinal herbs, including *Indigofera tinctoria, Angelica sinensis, Aloe vera, Moschus moschiferus, Phellodendron chinense, Saussurea lappa, Coptis chinensis, Gardenia jasminoides, Scutellaria baicalensis, Rheum palmatum*, and *Gentiana scabra* (*89*). *Qing Dai* (*Indigo naturalis*) among these 11 ingredients is the active constituent of the dark blue powder that is extracted from indigo dye-producing plants (*89*, *90*). *Qing Dai* is a mixture of the blue dye indigotin (**8**) and the red dye indirubin (**9**) (*89–91*). Indigotin (**8**) is inactive as the main component of *Qing Dai*, whereas indirubin (**9**) is a minor component with antitumor activity (*92*). Clinical studies for treatment of CML and chronic

granulocytic leukemia have shown the efficacy of indirubin (**9**), which led to 26% complete remission and 33% partial remission out of 314 cases (*89*, *91*). Low toxicity and minimal side effects have been detected in most of the patients evaluated (*89*, *91*). However, indirubin (**9**) has low water solubility and bioavailability, and some patients have suffered from severe gastrointestinal side effects (*91–93*).

5.1.2 Mechanism of Action

Cyclin-dependent kinases (CDKs), a family of serine/threonine kinases, are key regulators of the cell cycle and form enzyme complexes with regulatory subunit cyclins for CDK activity (*94–96*). Selective modulations of CDKs have been validated extensively as attractive molecular targets for developing human cancer therapeutic agents (*94–96*). Originally, indirubin was identified as a potent modulator of CDK1/cyclin B, CDK2/cyclin A, CDK2/cyclin E and CDK5/p25, thereby mediating strong growth inhibition of human tumor cells (*89*, *91*, *97*). The analysis of the cell cycle suggests indirubin (**9**) arrests tumor cells at the G_1/S or G_2/M phases (Fig. 12), resulting in cell growth inhibition and eventually, induction of apoptosis (*89*, *91*).

In an X-ray crystallography study of the CDK/indirubin complex, indirubin (**9**) was shown to bind competitively to the ATP-binding pocket as an ATP mimic in the
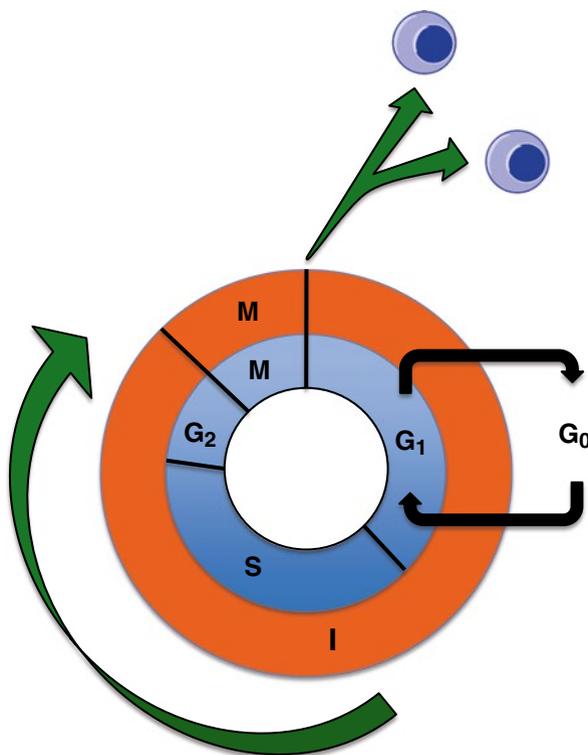
Fig. 12 Cell cycle

catalytic domain of CDKs. Three hydrogen bonds are formed in the ATP-binding pocket by the NH-CO-N′H functionality of indirubin and are crucial for binding to the peptide backbone of CDKs (*89*, *98*) (Figs. 13 and 14). Molecular modifications or substitutions of these groups of indirubin destroy the binding affinity to the ATP pocket of the catalytic domain and the ATP-competitive inhibition of CDKs.
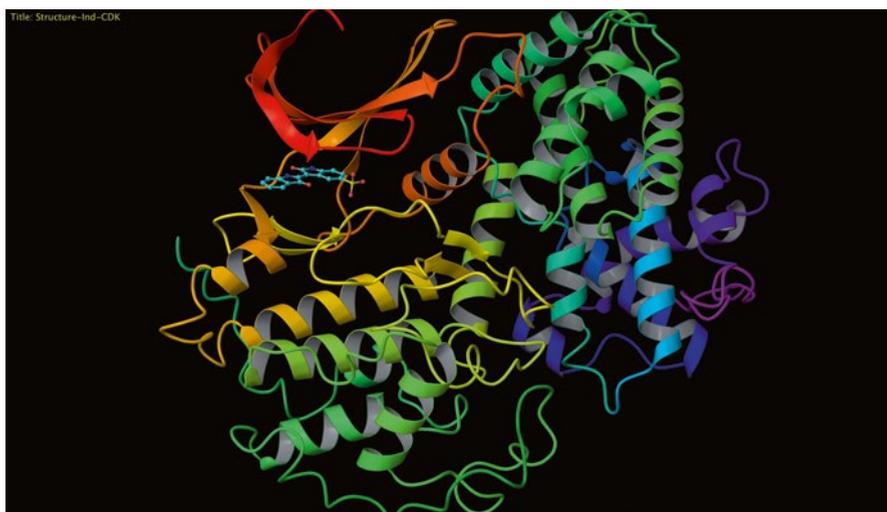
Fig. 13 Indirubin-5-sulfonate bound in CDK2. (The picture was created from the PDB file 1E9H using Maestro software, Academic version 9.3.5., Schrödinger Inc.)

Fig. 14 Indrubin-5-sulfonate in the ATP-binding pocket of CDK2. (The picture was created from the PDB file 1UV5 using Maestro software, Academic version 9.3.5., Schrödinger Inc.)

5.2 Royal Purple: A Source of Therapeutic Agents

5.2.1 Discovery of the Potential of 6-Bromoindirubin

The biological promises of the constituents of Tyrian Purple secreted from the hypobranchial glands of *Hexaplex trunculus* were identified using a bioactivity-guided isolation method (*62*). The inhibitory potential of each extract and fraction towards a selected panel of protein kinases (GSK-3, CDK1/cyclin B and CDK5/p35) has been evaluated. 6-Bromoindirubin (**13**) (6-BI) was first isolated and recognized as a selective inhibitor of glycogen synthase kinase-3β (GSK-3β) (*62*, *99*) (IC_{50} of 0.5 μM).

5.2.2 GSK-3β

Glycogen synthase kinase-3 (GSK-3) is a serine/threonine protein kinase that consists of two isoforms (GSK-3α and GSK-3β) with 97% homologous similarity (*100*). Originally, GSK-3 was discovered for its involvement in glycogen metabolism as the enzyme responsible for phosphorylation of glycogen synthase. It is the last link in the chain of phosphorylating agents triggered by the insulin-signaling pathway. The release of insulin inhibits GSK-3 signaling, leading to the storage of glucose (*101*, *102*). Due to its preponderant presence in the brain, GSK-3 has further been pinpointed as a valuable target to treat neurodegenerative disorders, including *Alzheimer*'s disease by preventing over-phosphorylation of Tau protein (*103–105*), mood disorders (*106*) and schizophrenia (*107*). In addition, GSK-3 was identified as a key intermediate of several important biological processes such as the Wnt (*100*) and Hedgehog (Hh) (*108*) signaling pathways. Canonical Wnt signaling controls differentiation at the stages of embryonic and neuronal developments, the fate of stem cells, and neuroprotection (*109*) and is deregulated in cancer (*110*, *111*). As for Wnt, Hh signaling is a complex pathway regulating embryonic development and tissue repair by controlling stem and progenitor cells (*112*). It is also deregulated in cancer (*113*). Therefore, GSK-3 appears to be an attractive target for cancer. Recently, GSK-3 activity has been proved to regulate also STAT signaling (*114*).

5.2.3 6BIO, a Biological Tool

Based on the observed results of 6-BI, a slight modification has been performed at the 3′-position and has led to the semi-synthetic analog, 6-bromoindirubin-3′-oxime (**80**) (6BIO, Fig. 15) (*115*).

(2Z,3′E)-6-Bromo-3′-(hydroxyimino)-(2′,3-biindolonylidene)-2-one, 6BIO (**80**) possesses an affinity 100-fold more potent than **13** towards GSK-3β ($IC_{50} = 5$ nM) (*62*) and a selectivity ratio CDK1/GSK-3 of 1/64 and CDK5/GSK-3 of 1/17 (*62*, *115*). The crystal structure of 6BIO (*62*) (Fig. 16, PDB 1UV5) inside the ATP pocket of GSK-3 revealed the binding mode and lights up as a reason for such great affinity.
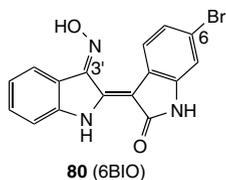
Fig. 15 Structure of 6BIO

Fig. 16 6BIO bound in the ATP-binding pocket of GSK-3. (The picture was created from the PDB file 1UV5 using Maestro software, Academic version 9.3.5., Schrödinger Inc.)

The presence of the successive NH-CO-NH pattern of the H-shape part of 6BIO (**80**) is the key to ensure the strong anchoring of the molecule by creating three hydrogen bonds inside the cavity with the two critical residues of Val135 and Asp133 (Fig. 17). The oxime at the 3′-position assures the completion of the binding by interacting with amino side chains through a water molecule (*115*). Moreover, the presence of Leu132 above the molecule provides enough space and hydrophobicity for the bromine to interact through *van der Waal's* forces. Indeed, the cavities of CDK2 and CDK5 contain a phenylalanine residue at the corresponding region of Leu132 thereby hindering the access of 6BIO (**80**) in their cavity (*62*). The perfect combination of key substituents on the indirubin skeleton and the shape of the protein residues confer the high selectivity of 6BIO (**80**) for the inhibition of GSK-3.

Due to the key role of GSK-3 in biological processes and its overregulation in diseases and the potent inhibitory activity of 6BIO (**80**), the latter exerts a broad

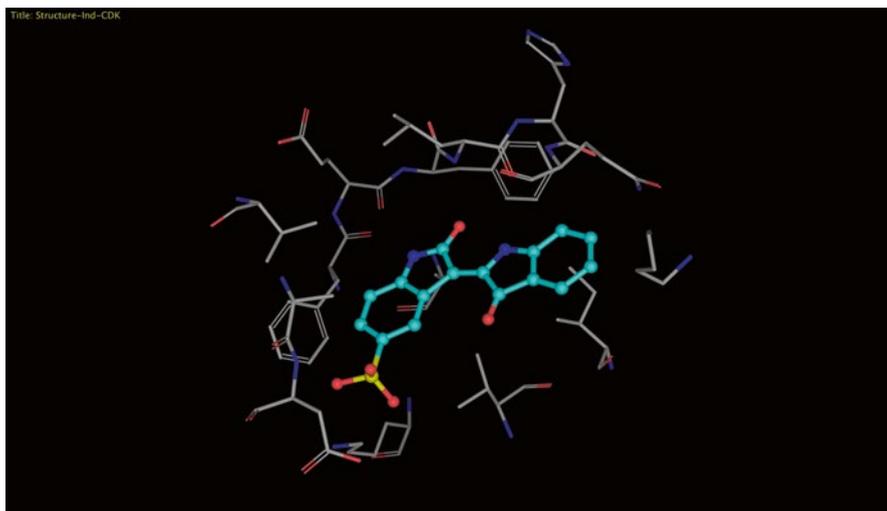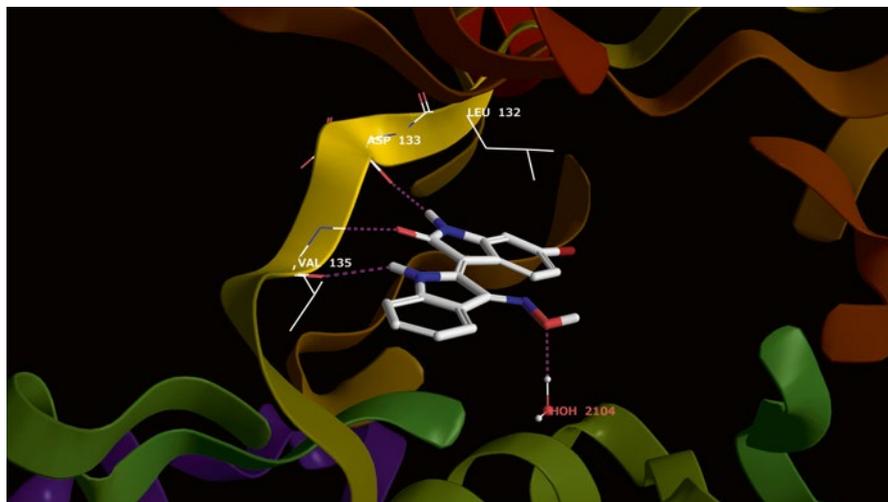**Fig. 17** Interactions of 6BIO inside the ATP-binding pocket of GSK-3. The H-bonds are shown as dotted lines. (The picture was created from the PDB file 1UV5 using Maestro software, Academic version 9.3.5., Schrödinger Inc.)

range of biological activities (*116*, *117*). However, discussed are only the most salient highlights.

5.2.3.1 Maintenance of Stem Cell Pluripotency

Embryonic stem cells are capable of self-renewal and differentiation in any type of cells (pluripotent cells) or into specialized cells. The Wnt pathway is notably the cellular signaling pathway controlling their fates. Indeed, numbers of its components were detected in human embryonic stem cells (*118*). The activation of Wnt needs GSK-3 to be switched off (*109*). Therefore, it has been proved that administration of 6BIO (**80**) to mouse and human embryonic stem cells was able to maintain their pluripotency (*119*). This important finding not only offers the possibility of creating sustainable *in vivo* models (*120*), but also opens the door to practical applications of regenerative medicine.

5.2.3.2 Antiparasitic Activity

As with all living organisms, parasites (*Leishmania* spp., *Trypanosoma* spp.) possess a kinome, although this is less developed than in the human body. The lower numbers of protein kinases present in these species play an important role in cell

cycle progression and survival (*121*). For a few years, studies have been engaged to identify parasitic protein kinases as biological targets (*122*), for the development of potent therapeutic agents that are currently lacking to treat neglected tropical diseases (NTDs) (*123*). The structure of parasitic GSK-3 has been determined (*124*) and is marked as a valuable target (*125*) along with other kinases like CRK-3 (*126*), the parasitic parent of human CDKs. A screening of kinase inhibitors has revealed that indirubin derivatives possess interesting antiparasitic activity (*127*). Due to its demonstrated ability to interact with the ATP pocket of kinases, subsequently 6BIO (**80**) has later been evaluated in *Leishmania donovani* and identified as a potent CRK-3 inhibitor (*128*). As a result, 6BIO (**80**) was one of the most potent parasitic kinase inhibitors found and this has propelled indirubins as useful scaffold molecules for the study of NTDs.

5.2.3.3 Anticancer Activity

Among the 518 kinases composing the human kinome, more than 80% are serine/threonine kinases (*129*). Moreover, it is known than the ATP-site where indirubins bind is well conserved all over the kinome (*129*, *130*). Investigations have shown that 6BIO (**80**) also inhibits several other kinases (*131*). Among its other targets, 6BIO (**80**) inhibited activities of JAKs *in vitro* and down-regulated constitutive activation of Stat3 signaling, associated with induction of apoptosis and suppression of tumor growth *in vivo*, in a human melanoma model study (*132*). In addition to these pharmacological effects on melanoma, 6BIO (**80**) induces apoptosis of aggressive breast tumor cells (*133*).

5.2.4 Novel Analogs of 6BIO

Novel 6BIO analogs containing extended amino-aliphatic chains on the 3′-oxime group have been designed *in silico* to accentuate the ATP mimicry. Their synthesis consists of two successive nucleophilic substitutions on the oxime (*134*). In a primary step, dibromoethane undergoes a first nucleophilic substitution by attack of the oxime in the presence of Et₃N to form an intermediate (**81**). A second step involves the removal of the second bromine by the corresponding amines (Scheme 25). This versatile synthesis method then allowed the creation of a new assembly of 40 derivatives (*134*).

Enzyme inhibition assays showed substantial improvement of both potency and selectivity of the new analogs against GSK-3*β*. Furthermore, the secondary or tertiary amines could be converted to their corresponding salts (by addition of HCl/Et₂O, Scheme 25) in order to enhance water solubility (*134*). These analogs exhibit strong antitumor activities against SH-SY5Y neuroblastoma cells (*134*) and a recent report highlights them as promising anti-*Alzheimer*'s agents by preventing phosphorylation of tau protein (*135*).
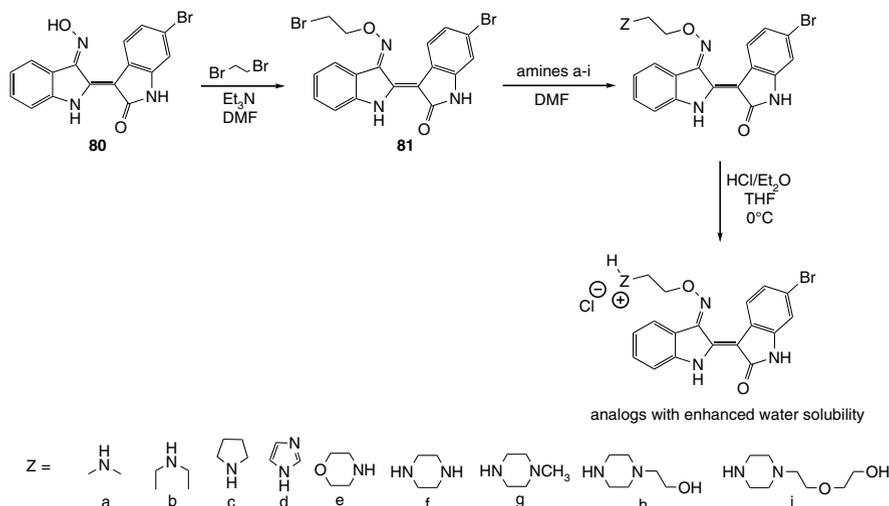
Scheme 25 Synthesis of water-soluble analogs of 6BIO

5.3 Indirubin Synthesis

As indirubin (**9**) was not chemically identified until the end of the nineteenth century and was not of equal importance for industrial development as indigotin (**8**), the development of methods for its synthesis was not a priority. However, as a natural by-product of indigo, indirubin (**9**) was first observed as a by-product formed during the synthesis of indigotin (**8**). Although *Schunk* had characterized the structure of this by-product and denominated it as "indirubin" (*136*), the synthesis of this bis-indole has required rather challenging chemistry to perform.

5.3.1 *Baeyer* Breakthrough

Baeyer first produced indirubin (**9**) by reduction of isatin (**27**) in 2-chloro-3*H*-indole-3-one (**82**), using phosphorus pentachloride (*39*, *40*, *137*) (Scheme 26). At that time, *Baeyer* noticed the formation of indigotin (**8**) in a low yield along with a red-colored product partially soluble in ethanol but not in water, which he called "indigo purpurin".

A few years later, *Baeyer* carried out the coupling between indoxyl (**26**) and isatin (**27**) (*138*) and found the first access to the indirubin core, following therefore the biosynthesis pathway (Scheme 27). Nonetheless, indoxyl (**26**) was a difficult intermediate to synthesize as it exhibits considerable instability and dimerizes quickly to indigotin (**8**).

The significant advance that represented the two methodologies (activation of the 2-position of isatin and biomimetic pathways) developed by *Baeyer* laid the foundations for the development of more efficient pathways.
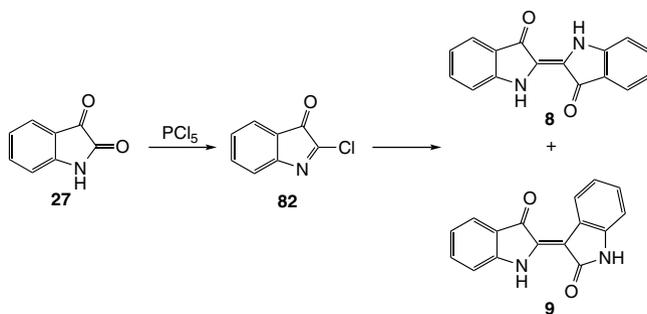
Scheme 26 First non-selective formation of indirubin (**9**)
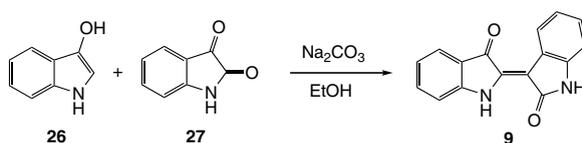
Scheme 27 First synthesis of indirubin (**9**)

5.3.2 Activation of the 2-Position

5.3.2.1 The 2-Chloro-3*H*-indole-3-one Pathway

Although the first formation of indirubin (**9**) could be categorized as a "happy accident", the reduction of isatin (**27**) by phosphorus pentachloride has nonetheless inspired chemists, who concentrated efforts towards the development of improved methodology. Indeed, *Wahl* (*139*) initially and *Katritzky* (*140*) later on successfully managed to control the formation of 2-chloro-3*H*-indol-3-one (**82**) from isatin (**27**). The intermediate was reacted directly with oxindole (**29**) in anhydrous pyridine (*139*) or toluene (*140*) to afford indirubin (**9**) in moderate to good yields (Scheme 28).
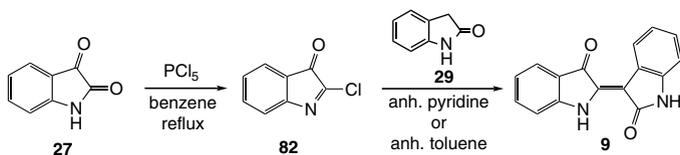
Scheme 28 Formation of indirubin through 2-chloro-3*H*-indol-3-one (**82**)

5.3.2.2 Stabilization of the Intermediate

Although the activation of isatin (**27**) to 2-chloro-3*H*-indol-3-one (**82**) has been performed with relative success, the aforementioned intermediate is quite unstable and has to be used readily after its formation under drastic anhydrous conditions. Several attempts therefore have then been made in order to generate stabilized activated isatin (**27**). *Wahl* and collaborators have carried out the synthesis of a new type of intermediate through the formation of the corresponding silver salt (**83**). The latter is reacted with methyl iodide to furnish a stable 2-methoxy-3*H*-indol-3-one (**84**) (*141*). Finally, **84** is then condensed with oxindole (**29**) in acetic anhydride and dry benzene in the presence of a catalytic amount of sulfuric acid, to furnish indirubin (**9**) (Scheme 29). A second stabilization method later developed by *Martinet* consisted in the formation of a 2-phenyliminoisatin (**85**) (*142*) (Scheme 29). The intermediate is then reacted with isatin (**27**) to afford indirubin (**9**) under an atmosphere of H_2S and in the presence of ammonia.
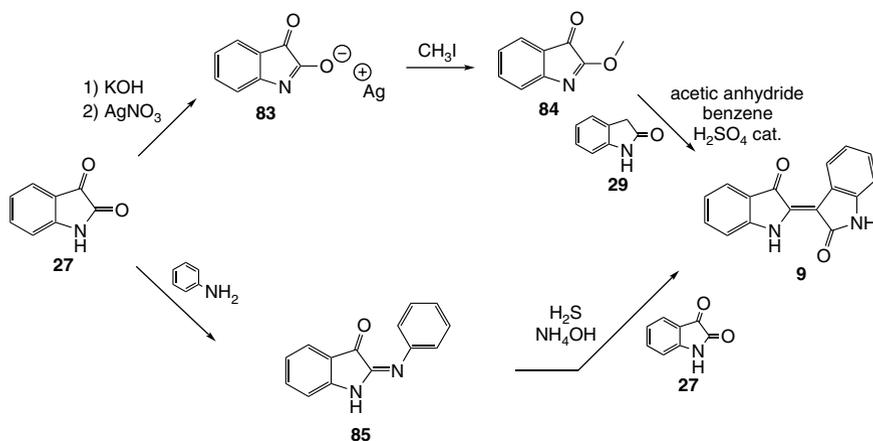
Scheme 29 Activation pathways of the 2-position

Later *Albert* defined a strategy involving the formation of a thioamide (**86**). The latter dimerizes to form indirubin (**9**) (65% yield) in the presence of sodium sulfite and sodium bicarbonate (*143*) (Scheme 30).
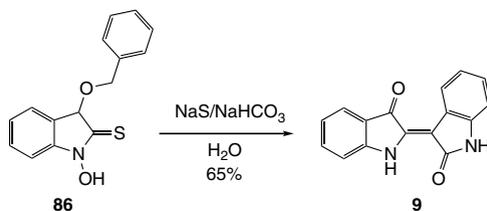
Scheme 30 Activation in thioamide (**86**)

5.3.2.3 Other Isatin Derivatives

The use of isatin (**27**) to study the formation of indigoids has been of great value, and *Ainley* and *Robinson* in their efforts to synthesize epindoline derivatives proposed an elegant alternative to obtain indirubin (**9**) (*144*) (Scheme 31). The reaction between isatin (**27**) and *o*-nitrophenacyl bromide (**87**) in alkaline medium (EtONa) affords isatylidene-*o*-nitroacetophenone (**88**), probably by rearrangement of *N*-*o*-nitrophenacylisatin (**89**) under basic conditions. The isatylidene-*o*-nitroacetophenone (**88**) is then converted to indirubin (**9**) in an EtOH-ammonia solution or by reduction of the starting material using zinc dust. Despite the efforts of the authors, the mechanism of formation is not yet clearly understood.
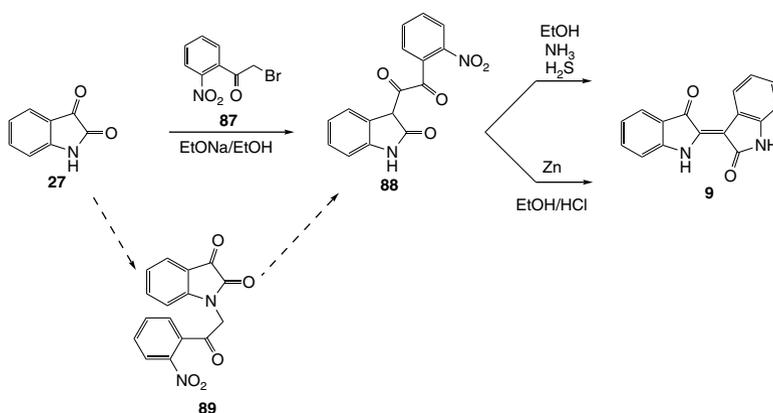
Scheme 31 Methodology involving isatylidene-*o*-nitroacetophenone (**88**)

Finally, a further method involves the reduction of isatin-3-oxime (**90**) with lithium aluminum hydride to form indirubin (**9**) (*145*) (44% yield), with the starting material being obtained from the condensation of isatin (**27**) and hydroxylamine (Scheme 32).
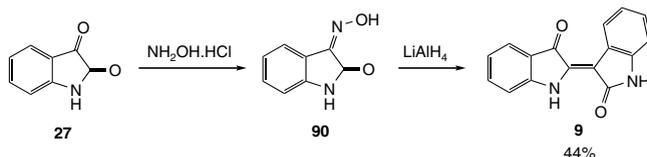
Scheme 32 Reduction of isatin-3-oxime (**90**)

5.3.3 Indoxyl Pathway

The biomimetic synthesis developed by *Baeyer* from indoxyl (**26**) indisputably has represented a major breakthrough and has been used by several researchers after its discovery (*136*, *146–148*). However, the ability of indoxyl (**26**) to dimerize to yield indigotin (**8**) has led to the necessity of developing a more stable intermediate.

5.3.3.1 Stabilization at the 2-Position

Martinet in his studies on indole derivatives proposed an alternative to the iminoisatin previously developed. Indeed, he used the reactivity of the 2-position of indoxyl (**26**) to condense the latter with phenylhydrazine (**91**) to obtain a stable 2-phenylazoindoxyl (**92**) (*149*). The latter is then reduced under basic conditions and in the presence of sodium sulfite to afford indirubin (**9**) along with ammonia and aniline (Scheme 33).
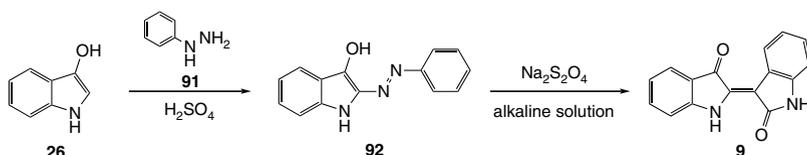
Scheme 33 Stabilized 2-substituted indoxyl

5.3.3.2 Indoxyl Acetate: The Ultimate Intermediate

During the same period, *Vorländer* achieved the second major breakthrough by stabilizing the indoxyl with the synthesis of indoxyl diacetate on the one hand and indoxyl acetate (**93**) on the other (*150*). The latter is obtained from the esterification of indoxyl salt (**94**) in a degassed aqueous solution of acetic anhydride (Scheme 34).
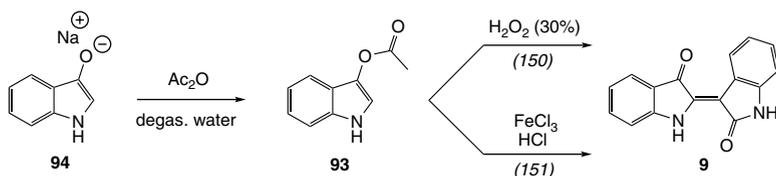
Scheme 34 Oxidation of indoxyl acetate

Indoxyl acetate (**93**) is then oxidized in the presence of hydrogen peroxide (30% solution) to form indirubin (**9**). This method inspired *Spencer*, who proposed an acidic hydrolysis using iron chloride (III) (*151*) (Scheme 34).

Finally, additional progress has been realized by *Russell* and *Kaupp* (*46*) by coupling indoxyl acetate (**93**) and isatin (**27**) in methanol in the presence of sodium carbonate. This pathway, as in *Baeyer*'s method, mimics the biosynthesis of indirubin (**9**) and has later been used widely for the generation of assemblies of derivatives (Scheme 35).
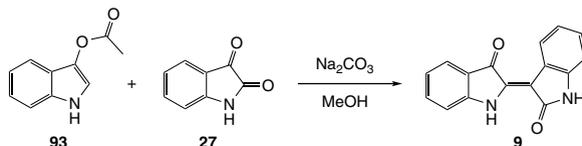
Scheme 35 Coupling between indoxyl acetate and isatin by *Russell* and *Kaupp*

5.3.4 Biotransformations

Indigoids, including indirubins, have been obtained by feeding *Polygonum tinctorium* tissue cultures with indoles (*152*). Although this method was useful for elucidation of the biosynthesis mechanism of indigotin (**8**) and indirubin (**9**), its application for synthesis is, however, limited.

Considering the involvement of oxidative media in the natural generation of indigoids, the use of enzymes that oxidize indoles has come to the forefront. In this approach, mutant human cytochrome P450 (expressed as *Escherichia coli*-mediated oxidation has been performed using 5-methoxy- or 5-methylindole (*153*). This "eco-friendly" method has not only permitted the generation of indigotin (**8**) and indirubin (**9**), but it has provided also new analogs such as 5-methoxyindigo, 5′-methoxyindirubin, 5,5′-dimethoxyindigo, 5-methoxyindirubin, 5,5′-dimethoxyindirubin, 5-methylindirubin, and 5′-methylindirubin. This assembly was then evaluated for the inhibition of protein kinases (CDK1, CDK5, and GSK-3) and all the indirubin derivatives were more active than their parent compound, in terms of their preference for GSK-3. Despite its ecological benefits, this production procedure for indirubins suffers from a lack of selectivity.

Nevertheless, this enzymatic approach has inspired chemists, and a catalytic cobalt (II)-based oxidative degradation method was developed for indoles, leading to the selective formation of indigoids (Scheme 36) (*154*). Indeed, it is reported that the mechanism of oxidation by cytochrome P450 resembles chemical oxidation by hypervalent transition metal oxidants. The Co(II) complex was first prepared using glycine and 8-hydroxyquinoline as ligands (*155*) followed by the successive addition of indole and hydrogen peroxide. The decomposition of H_2O_2 generates the oxidative species (O_2) expected to be involved in the study. The mixed ligand Co(II) complex and H_2O_2 led to the production of isatin (**27**), indigotin (**8**), and indirubin (**9**) in high yields and to the recovery of the catalyst.
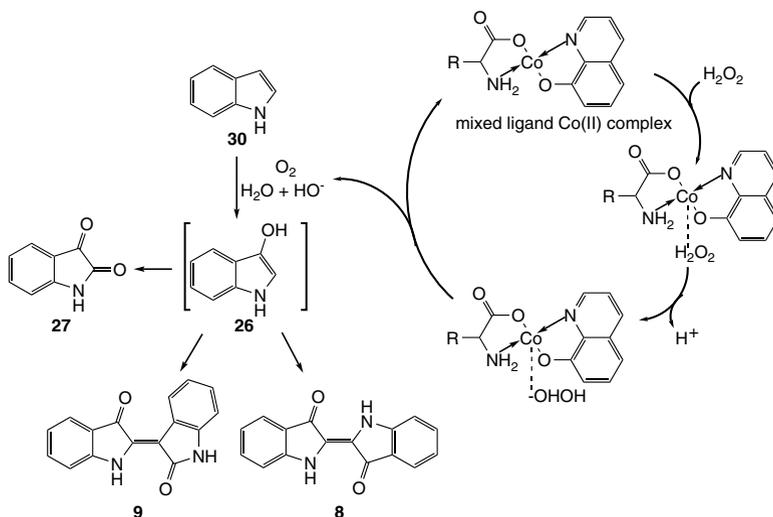
Scheme 36 Catalytic decomposition of hydrogen peroxide and application to indigoid synthesis

5.4 Biological Applications of Synthetic Indirubins

5.4.1 Creation of Compound Assemblies

Due to the demonstrated importance of indirubins for biological studies and drug development, the necessity of developing large numbers of derivatives has become evident, in order to conduct structure-activity relationship studies (SAR). For quick and easy access, the methodology of *Russell* and *Kaupp* has been widely applied as it offers the maximum flexibility for the introduction of new substituents. Therefore, the dimerization of a correctly substituted isatin and the commercially available 3-acetoxyindole can lead to the synthesis of substituted indirubins. In this manner, a few indirubin analogs (*89*, *93*, *98*) and 30 new indirubins bearing chlorine, fluorine, iodine, or vinyl groups on the 6-position were obtained, but 5,6-disubstituted derivatives were synthesized initially (*115*). The selection of 5- or 7-substituted isatins opened the door to the creation of new libraries of synthetic analogs possessing diverse substituents at the 5- and 7-positions (*156–160*). Interestingly, the versatility of the method is conserved when starting from azaisatin or 3-acetoxy-5- or 6-substituted-indoles, leading, respectively, to azaindirubins (*161*, *162*) and di-substituted indirubins (*158*, *163*, *164*). Overall, the general synthesis scheme can be represented as depicted in Scheme 37. Owing to this versatile synthesis method, the pharmacological benefits of indirubins have been explored rapidly as a result of this production method for promising indirubin analogs.
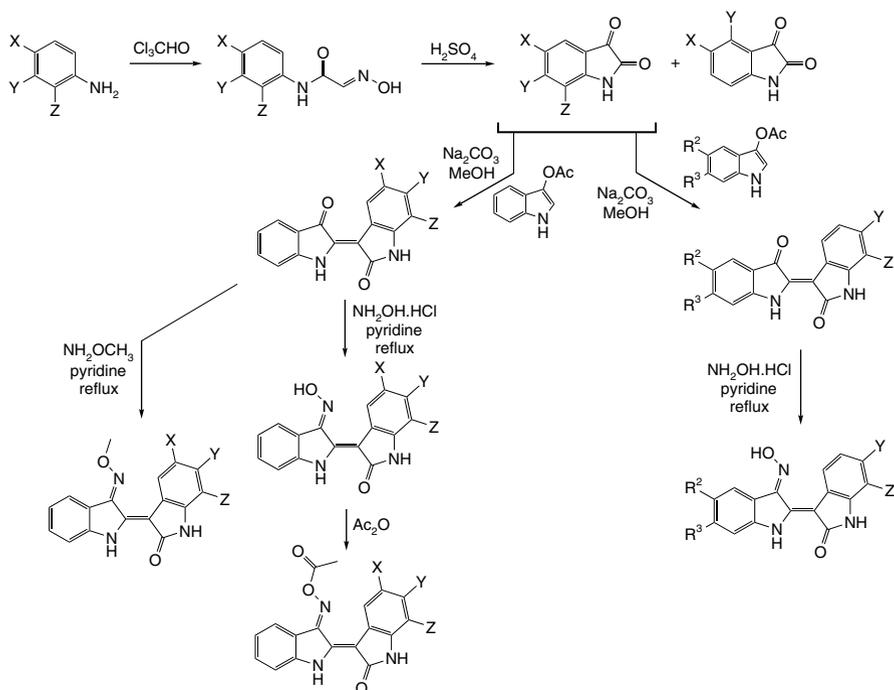
Scheme 37 General scheme of the synthesis of indirubin analogs

5.4.2 Indirubin Derivatives

The 5- or 3′-positions of indirubin (**9**) can be modified for improving potency, selectivity, and solubility (*93*, *98*). These chemically modified or substituted small molecules have been actively evaluated for their cytotoxic and antitumor activities using systems of purified enzymes *in vitro*, cell cultures, and *in vivo* models (*89*). Among synthetic indirubin analogs, indirubin-3′-monoxime (**94**) (IO, Fig. 23) and indirubin-5-sulfonate (**95**) (Fig. 18) show potent inhibitory activities against CDKs, and induce cell growth inhibition of human MCF-7 breast cancer cells (*91*, *92*).

Several different analogs were synthesized based on the reactivity of the oxime. Indeed, under basic conditions, diverse side chains can be introduced at the 3′-position (*e.g.* glycerol or glucose) leading to the synthesis of derivative E804 (**96**) (Fig. 18) (*93*). Concerning the biological mechanism of action, a number of indirubin analogs potently block constitutively activated Stat3 signaling of human breast and prostate cancer cells. Particularly, *in vitro* biochemical kinase assay data suggest E804 containing a dihydroxypropyl 3′-oxime ether substituent as a potent inhibitor of Src kinase, down-regulating constitutively activated Stat3 or Stat5 in human breast cancer, CML, and T315I mutant Bcr-Abl expressing CML cells (*93*, *165*). However, prolonged modulation against Src activity can cause a recovery of impaired Stat3 activation and thus tumor cell survival through altered JAK/Stat3 interaction (*166*, *167*).
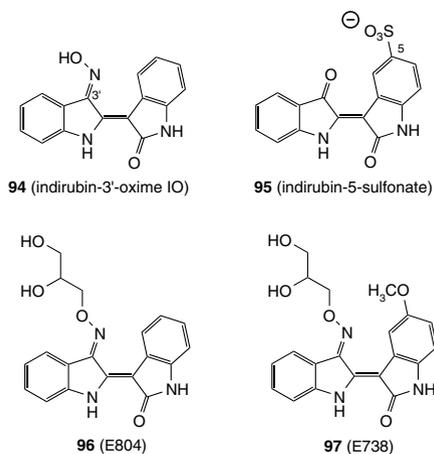
Fig. 18 Structures of indirubin analogs

This compensatory feedback implies that dual inhibition of JAKs and the Src family of kinases (SFKs) is probably a more favorable molecular target to inactivate downstream Stat3 signaling in cancer therapy (*107*, *168*). On the basis of the previous findings, indirubin analog E738 (Fig. 18) was synthesized mainly to improve water solubility and bioavailability (*91*, *168*). Unexpectedly, the small molecule E738 (**97**) appears to be a strong dual inhibitor of JAKs/Stat3 or SFKs/Stat3 signaling, associated with induction of apoptosis of human pancreatic cancer cells (*107*, *168*).

5.4.3 7-Bromo-Indirubins

Isomers of 6BIO (**80**) have been developed with the introduction of bromine and other halogens in the 7-position (*159*). As for 6BIO, 7BIO (**98**) exerts a potent cytotoxic activity (*131*, *159*) (Fig. 19). Surprisingly, the displacement of the
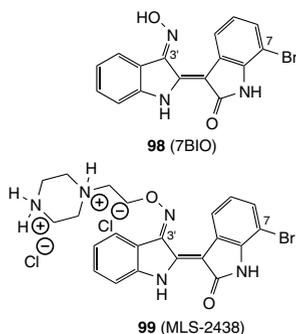
Fig. 19 7-Bromoindirubins

bromine on the 7-position induces a shift in kinase selectivity. Indeed, 7-bromoindirubins have been proved to be potent Aurora B and C inhibitors (*169*). The aforementioned kinases are intimately involved in cell cycle regulation and are valuable biological targets for cancer cell killing. Furthermore, *in vivo* evidence suggests a water-soluble analog of 7-BIO (**99**), MLS-2438 (**99**) (Fig. 19), can suppress the survival of human melanoma cells associated with inhibition of Stat3 and Akt signaling.

5.4.4 New DYRK 1A Inhibitors

The protein kinase DYRK1A is encoded by the *DYRK1A* gene localized on chromosome 21. It is involved in cell cycle progression and neurodegenerative diseases such as *Down* syndrome. This kinase has been poorly studied because of the lack of potent inhibitors. The promising scaffold provided by 7-BIO analogs resulted in the modification of the ring leading to the development of 7-bromo-5′- or 6′-substituted indirubins (*163*). Carboxylic acid, formyl, oxime, ethanol, and cyano groups have been introduced with success at the above-mentioned positions. Interestingly, the insertion of a tetrazole moiety (7-bromo-5′-tretazolindirubin (**100**)) as a bioisostere of the carboxylic acid has been carried out from the corresponding 7-bromo-5′-cyanoindirubin (**101**) using the stable trimethylsilyl azide (Scheme 38). To the present, this molecule is the first and only example of the introduction of an extra ring on the indirubin skeleton.
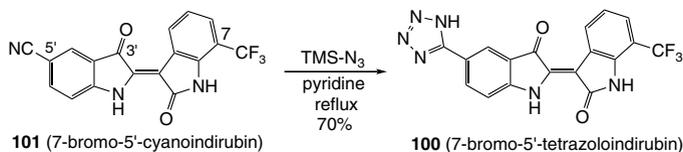
Scheme 38 Synthesis of the 7-bromo-5′-tetrazoloindirubin (**100**)

Moreover, a new series of 7-trifluoromethylindirubins has also been developed for biological investigation (Fig. 20). Interestingly, the presence of chemical entities on the positions 5′- and 6′- confers another type of kinase inhibition to the new analogs. In fact, di-substituted 7-BIO indirubins are potent and selective DYRK kinase inhibitors, and 7-bromo-3′-oximindirubin-5′-carboxylic acid (7-BIO-5′-COOH (**102**)) (Fig. 20) is the lead compound for this new series of inhibitors (*163*).

Interestingly, the binding mode of 7-BIO-5′-COOH was unexpected as the indirubin binds in an inverse fashion when compared to traditional anchoring (*163*) (Fig. 21). In this specific case, the bromine was exposed to both the solvent and the carboxylate group formed a salt bridge with the lysine 178 residue inside the cavity.

R = COOMe, COOH,
CN, tetrazole

7-trifluoromethylindirubins

102 (7-BIO-5'-COOH)
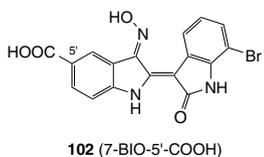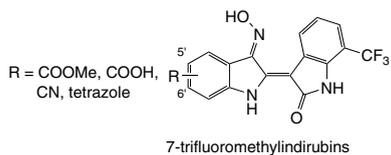
Fig. 20 New type of DYRKs inhibitor
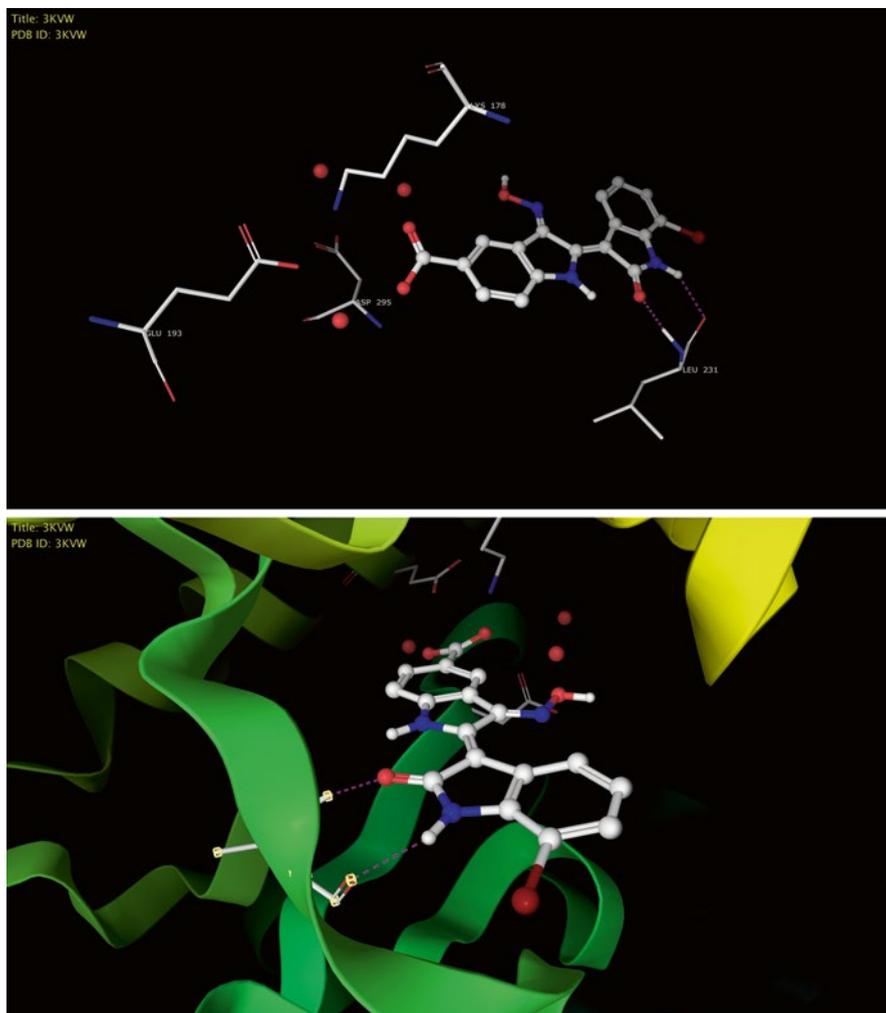
Fig. 21 Zoom on the interactions of **102** in the DYRK cavity. (The pictures have been created from PDB file 3KVW using Maestro software, Academic version 9.3.5., Schrödinger Inc.)

Moreover, the traditional NH-CO-NH pattern was not entirely involved with binding as only the lactam ring components formed a hydrogen-bond with leucine 231. Hence, the new indirubins represent valuable biological tools to deepen knowledge on the role and the impact of DYRK kinases in the biological system.

5.4.5   Other Indirubin Derivatives

Exploration of the 5-position was recently led to some success (*170*). Compounds such as 5-fluoro-indirubin-3′-oxime (**103**) (Fig. 22) have been recognized as potent FMS-like tyrosine kinase-3 (FLT3) kinase inhibitors (*171*) involved in cancer development and especially leukemia (*172*, *173*).

As another novel family of synthetic indirubins, 5,5′-disubstituted indirubin-3′-oxime analogs, such as 5-nitro-5′-hydroxyindirubin-3′-oxime (**104**) (Fig. 20), also exhibit potent inhibition against CDK2/cyclin E *in vitro*, and in addition, an induction of apoptosis and *in vivo* efficacy (*157*, *164*).
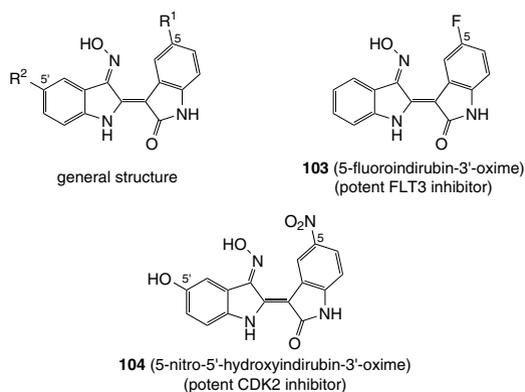
Fig. 22   5,5′-Disubstituted indirubins

5.4.6   Conclusion

Although indirubins are minor compounds of ancient dyes, they provide interesting scaffolds with several applications in biology. These unexpected applications have allowed the discovery of powerful ATP-mimetic/ATP-competitive serine-threonine or tyrosine kinase inhibitors. Moreover, indirubins are modulators of GSK-3 signaling, which plays a key role of the maintenance of stem cell pluripotency. These crucial breakthroughs may contribute to the development of regenerative medicine therapies. Chemists involved with the design of indirubin derivatives might well observe that "a trip around the indirubin scaffold is a trip around the human kinome" (Fig. 23).
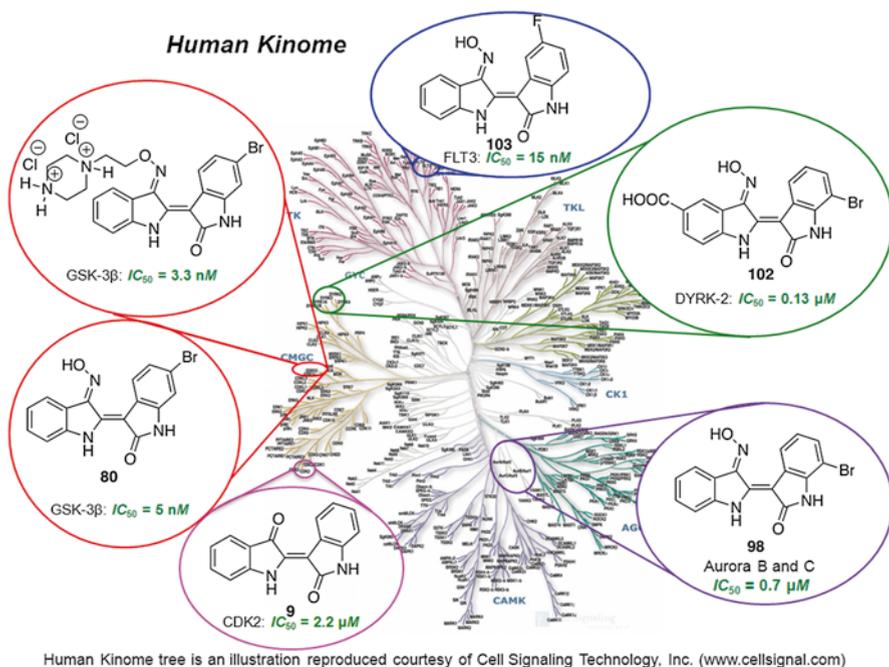
Fig. 23 Distribution of indirubin kinase selectivity around the human kinome

5.5 Glycoside Indigoids

5.5.1 Glycosides of Indigos

5.5.1.1 The Family of Akashins

As mentioned in the first part of this contribution, the indigoid family includes natural halogenated-*N*-glycoside indigos. Akashins A, B, and C (**20–22**) have been isolated from *Streptomyces* sp GW48/1497 (*12*). Akashins are the first examples of natural 5,5′-disubstituted indigoids and *N*-glycoside indigos (even if synthetic compounds are included). As observed for their non-substituted parent molecule, akashins have low solubility in non-polar solvents even if the presence of the glycosidic residue enhances the solubility in methanol (but not in water). Nevertheless, akashins A, B, and C (**20–22**) possess cytotoxic activities against various human cancer cell lines (*12*) (colon carcinoma, melanoma, lung carcinoma, breast cancer, kidney tumor). Thus, their structures represent the first indigotin-based scaffolds to have been utilized in medicinal chemistry.

5.5.1.2 Synthesis of Akashins Derivatives

The challenging chemistry and the novelty of the structures of akashins led chemists to explore the total synthesis of their derivatives. *Hein et al.* first proposed an approach by O-glycosylation of *N*-benzylindigo (*174*). However, the key step consisting of a O→N rearrangement offers much less flexibility as it depends on the nature of the carbohydrate moiety (*174*) and its protective group. Therefore, the synthesis of protected pivaloyl-*N*-rhamnosylindigo (**105**) was achieved (Scheme 39), but the deprotection of the sugar moiety was unsuccessful.
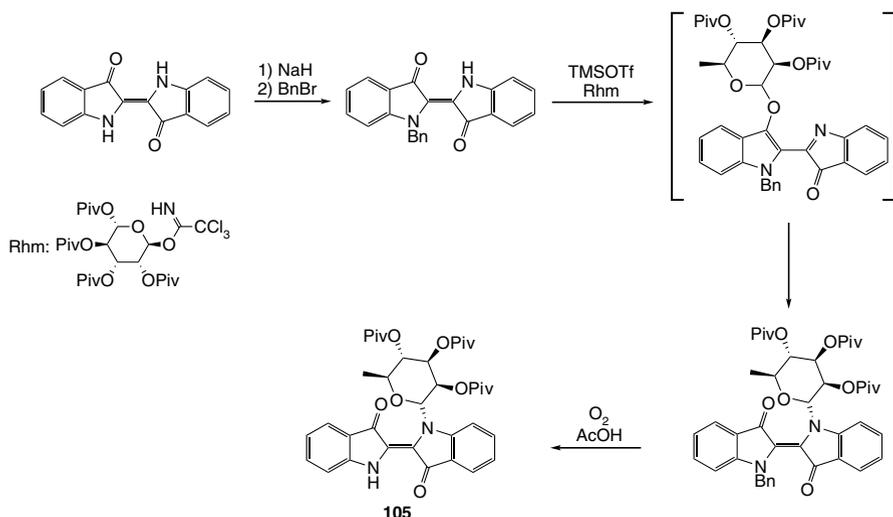
Scheme 39 Synthesis of protected *N*-pivaloylrhamnosylindigo (**105**)

Later, the same group proposed a novel approach using the reactivity of dehydroindigo (**106**) (*175*). The latter was prepared from indigotin (**8**) by the oxidative addition of acetate groups on carbons C-2 and C-9 following by elimination of acetic acid in an alkaline medium. The reaction of dehydroindigo (**106**) with TMS protected L-rhamnoside and subsequent acetolysis provided *N*-(2,3,4-tri-*O*-acetyl-L-rhamnosyl)indigo (**108**). The deprotection of the sugar in the presence of a catalytic amount of *t*-BuONa in methanol finally afforded *N*-rhamnosylindigo (**109**) (Scheme 40a). This sequence of reactions has then been applied successfully to mannosyl and glycosyl derivatives (Scheme 40b). However, the total synthesis of the natural akashins still represents an interesting challenge for chemists.
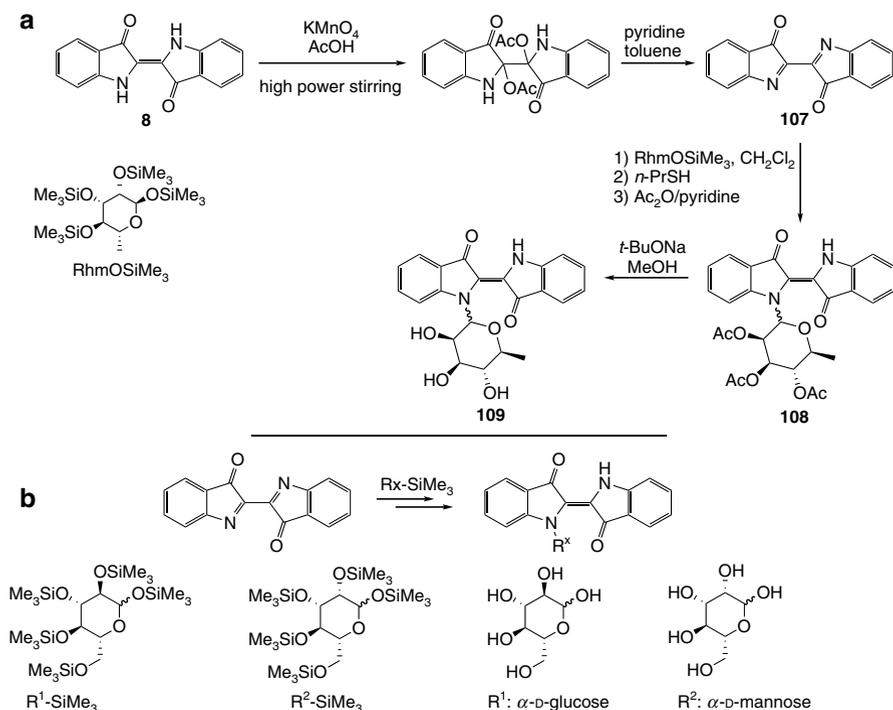
Scheme 40 Synthesis of deprotected indigo *N*-glycosides

5.5.2 Glycosides of Indirubins

In continuation of their efforts to obtain the akashins, *Libnow et al.* became involved in the synthesis of glycosidic indirubins.

5.5.2.1 Indirubin *N*-Glycosides

The investigators first proposed the synthesis of *N*-glycosylated indirubins (*176*), elegantly named as "red sugars", adopting the *Russell* and *Kaupp* methodology. The key step relies on the synthesis of a N-glycosylation from aniline. The first step consists of the introduction of aniline at the anomeric position of rhamnose (**110**). After protecting the sugar moiety, the aniline is condensed in isatin (**27**) through an aluminum chloride-mediated cyclization in the presence of oxalyl chloride to form *N*-(2,3,4-tri-*O*-acetyl-α,β-L-rhamnopyranosyl)isatin (**111**). A pure sample of the β-isomer **112** was then reacted with indoxyl acetate, using an excess of sodium carbonate to ensure the deprotection of the rhamnose, leading to the desired *N*-(β-L-rhamnopyranosyl)indirubin (**113**) in a good yield (77%) (Scheme 41).

This efficient methodology has opened the way for the creation of a small assembly of derivatives (**114–120**, Fig. 24), including *N*-glucosyl- (*e.g.* **117** and **118**),
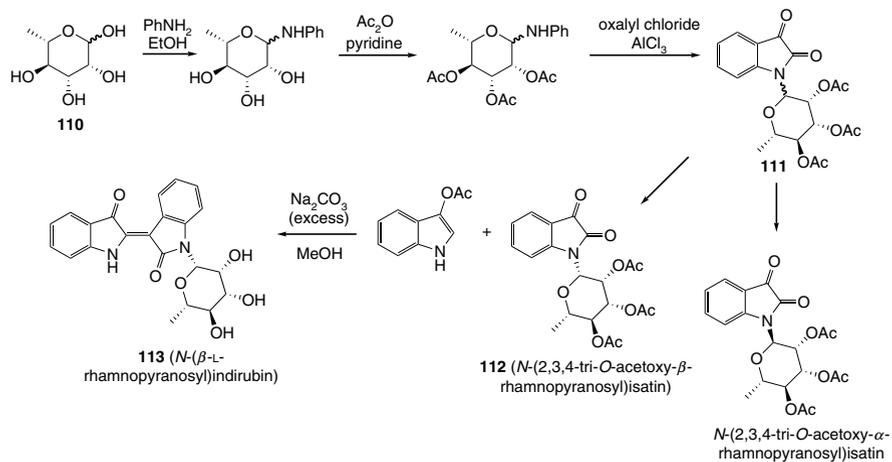
Scheme 41 Synthesis of the first *N*-rhamnosylindirubin (**113**)
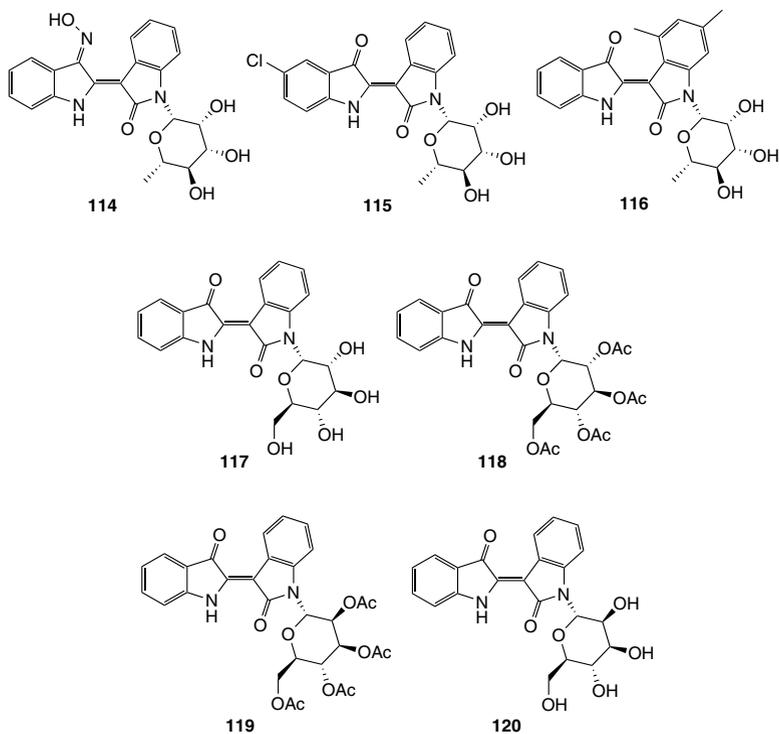
Fig. 24 Assembly of *N*-glycosylindirubin derivatives developed by *Libnow et al.*

N-galactosyl- (*e.g.* **119** and **120**), and *N*-mannosylindirubins (*177*). Moreover, some 5′-chloro-substituted indirubins (*e.g.* **115**), structurally closer to the akashins than previous analogs made, were also synthesized. The compounds were evaluated for their cytotoxic activities against four cell lines (5637, A-427, KYSE-70, MCF-7), giving the opportunity to perform a preliminary SAR analysis. Interestingly, the compounds exhibited a clear selectivity towards MCF-7 breast cancer cells. Furthermore, only the non-substituted *N*-rhamnosylindirubins exerted potent activity for MCT-5 cells, with IC_{50} values of between 0.67 and 0.76 μ*M*.

5.5.2.2 *N′*-Glycosylindirubins

In order to extend their study of glycosyl indirubins, *Libnow et al.* succeeded in obtaining *N′*-glycosylindirubins by developing the first synthesis of N-glycosylated indoxyls (*178*). For this purpose, indoline (**121**) was first reacted with rhamnose. 1-*N*-Rhamnosylindoline (**122**) was subjected to a DDQ-mediated dehydrogenation to provide the corresponding *N*-rhamnosylindole, which was directly protected (using BnBr (**123**) or MeI (**124**)). Next, the sugar-protected indole was used in the *Arnold* sequence consisting in the introduction of iodine in position-3 followed by its substitution using silver acetate to provide the protected *N*-rhamnosyl-3-acetoxyindole (**125–126**). The latter was condensed with isatin (**27**) using optimized conditions (piperidine, benzene) to form the protected *N′*-rhamnosylindirubin **127** and **128**. Indeed, the application of the classical conditions (methanol, Na$_2$CO$_3$) resulted in the formation of complex mixtures due to the instability of the indoxyl under such conditions. Finally, the deprotection of the sugar in a sealed tube (the steric hindrance reduces the use of less drastic conditions) afforded for the first time the corresponding *N′*-rhamnosylindirubin (**129**) (Scheme 42).
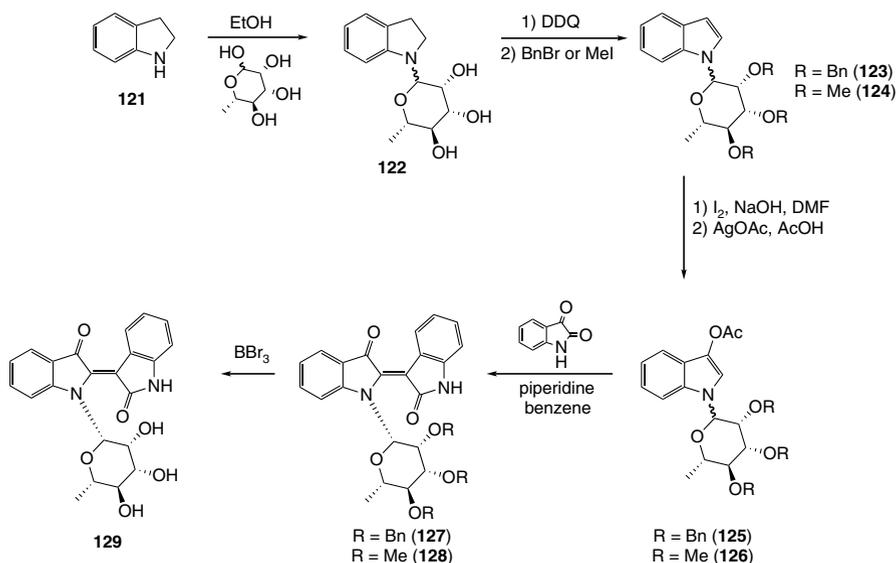
Scheme 42 Synthesis of *N′*-rhamnosylindirubin

5.5.2.3 *O*-Glycosylindirubins

Due to the positive biological attributes of indirubin (IO (**94**), especially), and as a result of its lack of solubility in biological media and slow metabolism in animals, the introduction of sugar moieties directly on the skeleton of this molecule is particularly attractive. However, the absence of substituents on nitrogen is critical for keeping the affinity of indirubin for the ATP-binding site of kinases. *Eisenbrand* and his group (*179*) were the first to explore the reactivity of the oxime at the 3′-position. Indirubin (**9**) was formed according to the method of *Russell* and *Kaupp* and IO (**94**) was obtained by action of hydroxylamine hydrochloride. Then, the coupling between acetylated *O*-chloroethylglucose (**130**) (prepared from an acetyl-protected glycosylfluoride (**131**)) and IO in basic conditions provided the corresponding monoxime ether containing an unprotected sugar moiety (**132**) (Scheme 43a). Another method (*179*) consists of a direct glycosylation of arabinose (**133**) using a haloalcohol and subsequent coupling reaction with IO (**94**) to form the corresponding indirubin-3′-(2-ʟ-arabinopyranosyloxyethyl) oxime (**134**) (Scheme 43b). The major advantage of this procedure is the avoidance of a multi-step protection-deprotection sequence.

Finally, in order to increase the rate of metabolism of indirubin, it has been proved that the introduction of a methyl or a methoxy at the 5-position can be a valuable option. Moreover, after administration of 5-methylindirubin (**135**)
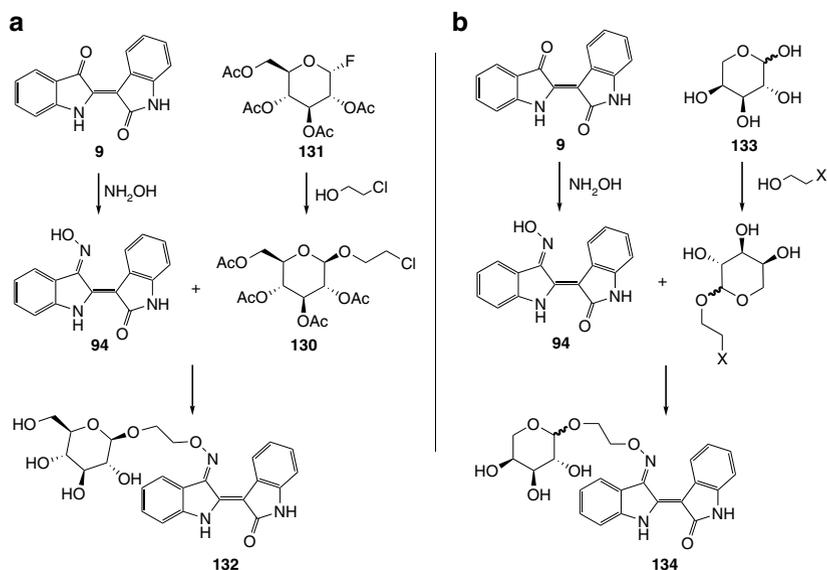
Scheme 43 Glycosylation of the 3′-position
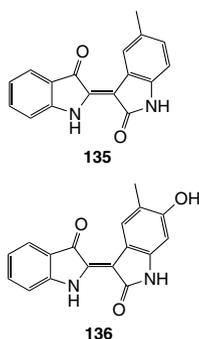
Fig. 25 5-Methyl- and 5-methyl-6-hydroxyindirubin

(Fig. 25), the first metabolite identified was 5-methyl-6-hydroxyindirubin (**136**) (Fig. 25), showing increased solubility and consequent overall interest from a chemical point of view.

Considering the importance of glucuronide derivatives for cell penetration and metabolism (the glucuronide would be released *in situ* by action of glucuronidase), the synthesis of 6-glucuronated indirubins was conducted (Scheme 44). Thus, 6-hydroxy-5-methylisatin (**137**) was synthesized initially from 5-amino-2-methylphenol (**138**). After a first step of phenol and amine protection (with triisopropylsilyl and Boc groups), the protected compound was engaged in a metallocyclization (directed *ortho*-metallation) using diethyloxalate. The compound 6-hydroxy-5-methylisatin (**137**) was finally recovered after acidic hydrolysis. This derivative was then involved in a coupling reaction with indoxyl acetate to
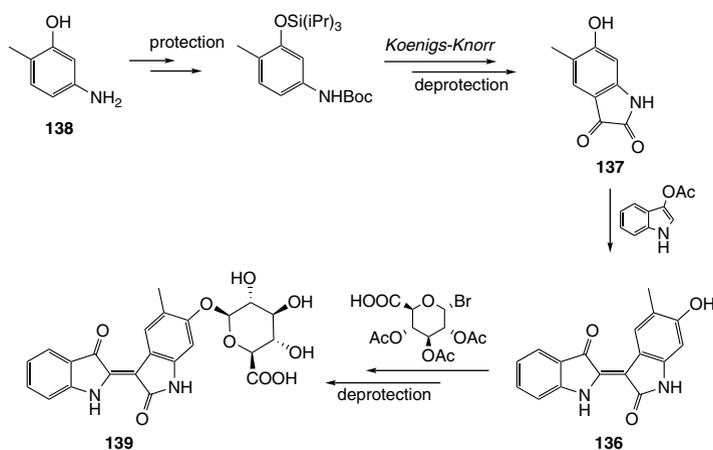
Scheme 44 Synthesis of 6-glucoronide-5-methylindirubin (**139**)

provide **136**. The glycosylation step under *Koenigs-Knorr* conditions followed with the deprotection of the glycoside, and finally afforded 6-glucuronide-5-methylindirubin (**139**). This synthesis strategy has been patented (*180*) and the compound claimed as a potential anticancer candidate.

5.6 Isoindigo: A Forgotten Family Member

Isoindigo (**11**) has been considered for a long time as a by-product that alters the quality of indigo dye. Manufacturers (traditional approaches or the chemical industry) have used their skills to eliminate isoindigo (**11**) from their preparations. The extensive interest of indirubin (**9**) for its biological applications has stimulated medicinal chemists to pay more attention to this relatively forgotten member of the indigoid compound class.

5.6.1 Synthesis of Isoindigo

As described previously, major efforts have been directed towards the synthesis of indigotin (**8**) and indirubin (**9**). The first sustainable synthesis approach for isoindigo (**11**) was developed much later, when compared to indigotin (**8**) and indirubin (**9**). *Wahl* and his collaborator *Bagard* first proposed a quantitative synthesis starting from oxindole (**29**) in the presence of 2-methoxy-indol-3-one under acidic conditions (*i.e.* acetic acid and a catalytic amount of hydrochloric acid) (*181*, *182*). This method was then improved with the replacement of 2-methoxy-indol-3-one by isatin (**27**) and now has been applied widely (*16*, *183–187*). It now represents the most well-known approach to the synthesis of isoindigo (**11**) (Scheme 45).
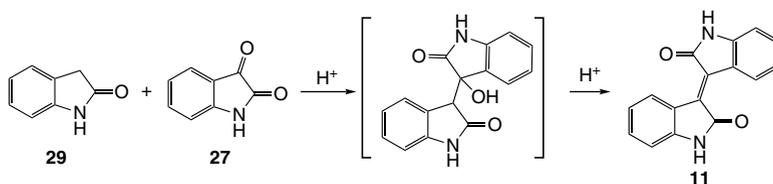
Scheme 45 Most applied synthesis of isoindigo (**11**)

Other investigators have proposed alternative pathways based on the activation of the 3-position of isatin with the introduction of diazomethane (*188–190*) or phosphine (*191*, *192*) (for a subsequent *Wittig* reaction) or oxindole (**29**) with sulfine (*193*, *194*) (using thionyl chloride), a disulfur bridge (3,3-oxindole dimers condensed under basic conditions) (*195–198*) and nitrososulfonate (*199*) (followed by acidic hydrolysis).

5.6.2 Isoindigo in Medicinal Chemistry

The side effects that were first observed after administration of natural indirubin (**9**), enticed medicinal chemists to develop new indigoid analogs, including isoindigo derivatives. However, as shown also for indirubin (**9**), isoindigo (**11**) has a low solubility in biological media, so the changes applied to its structure have been similar to those for indirubin (**9**).

5.6.2.1 Halogenated Isoindigo

The versatile chemical pathway developed some time ago by *Wahl* and *Balard* led to the synthesis of brominated isoindigo (*182*), but the investigators were not interested in evaluating biological activities at that time. Later on, several fluoro- (**140**), chloro-[(**141**) and (**142**)], and trifluoromethylisoindigo (**143–145**) (Fig. 26) analogs were synthesized and evaluated for their antiproliferative effects (*186*). The positive results obtained have highlighted the biological potential of this class of compounds, which were shown to be potent inducers of NAD(P)H-quinone oxidoreductase (NQO1).
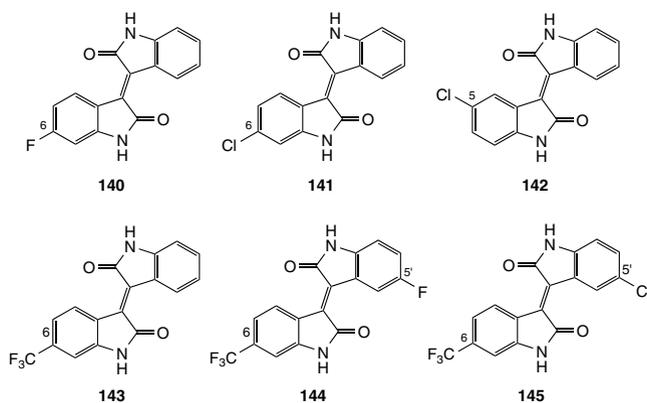
Fig. 26 Halogenated isoindigo inducers of NQO1

5.6.2.2 Glycosylisoindigo

In order to overcome the inherent solubility problems, glycosylisoindigo was developed. The lead compound in this series is Natura® (**146**). This *N*-glycosylisoindigo has been synthesized by *Wang et al.* in 2003 (*200*) from an acetyl-protected isatin *N*-glycoside. The latter is obtained by aluminum chloride-mediated ring closure from glycosyl aniline using oxalyl chloride (see Scheme 41). The reaction of *N*-glycosylisatin and with diverse oxindoles in acidic conditions converts the
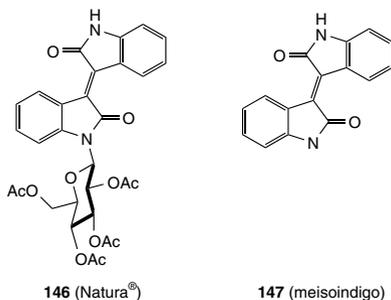
Fig. 27 Isoindigo derivatives

products into the desired isoindigo derivatives inclusive of *N*-methylated isoindigo
(**147**) (Meisoindigo). Natura® and meisoindigo (Fig. 27) exerted good cytotoxicity
towards various cancer cell lines. Interestingly, the acetyl-protected glycosylisoin-
digo was more active than its deprotected analog.

Based on those interesting results, *Sassateli et al.* have generated later an assem-
bly of diversely substituted derivatives of glycosylisoindigo (Fig. 28) and have
evaluated their antiproliferative activities (*201–203*).
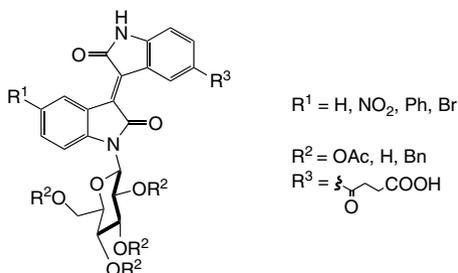
Fig. 28 General structure of the chemical assembly generated by *Sassateli et al.*

A final group of compounds in this class was synthesized as previously described
even though access to the N-glycosylation was redesigned. Starting from a substi-
tuted indoline (**148**), the investigators introduced the glucose moiety (**149**) and built
the isatin (**150**) by successive oxidation of the 2,3-positions (DDQ then CrO$_3$,
Scheme 46). The coupling between the aforementioned isatin (**150**) and oxindole
(**29**) followed by the deprotection of the sugar finally led to substituted isoindigo
derivatives (**151**). These derivatives were then evaluated for their cytotoxic potency
and exhibited moderate inhibitory activities.

The access to more soluble analogs has been a driving force in the medicinal
chemistry of indigoids. Therefore, the same team proposed the synthesis of the new
7′-azaisoindigo (Scheme 47), which was diversely substituted on both aromatic
parts of the molecule (Fig. 29) (*204*, *205*).
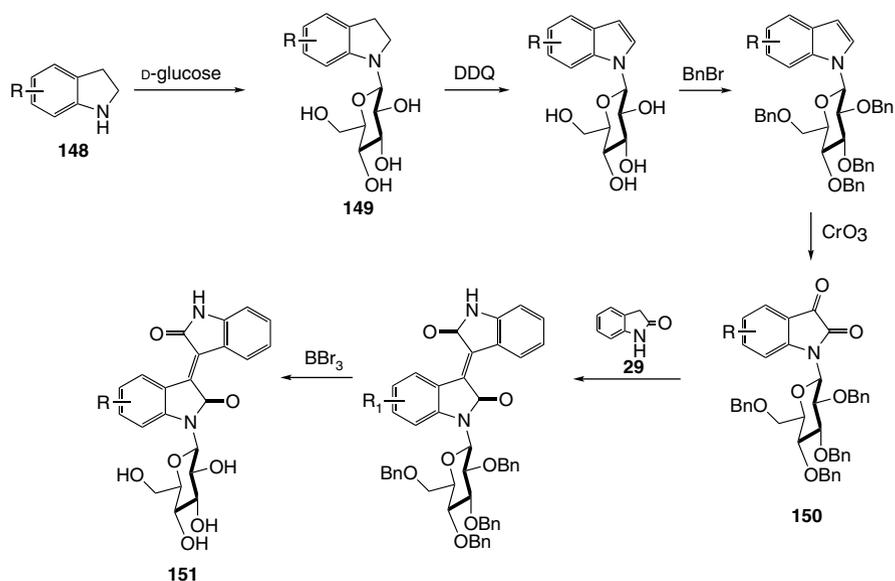
Scheme 46 Novel access to N-glycosylation applied to *N*-glycosylisoindigo synthesis
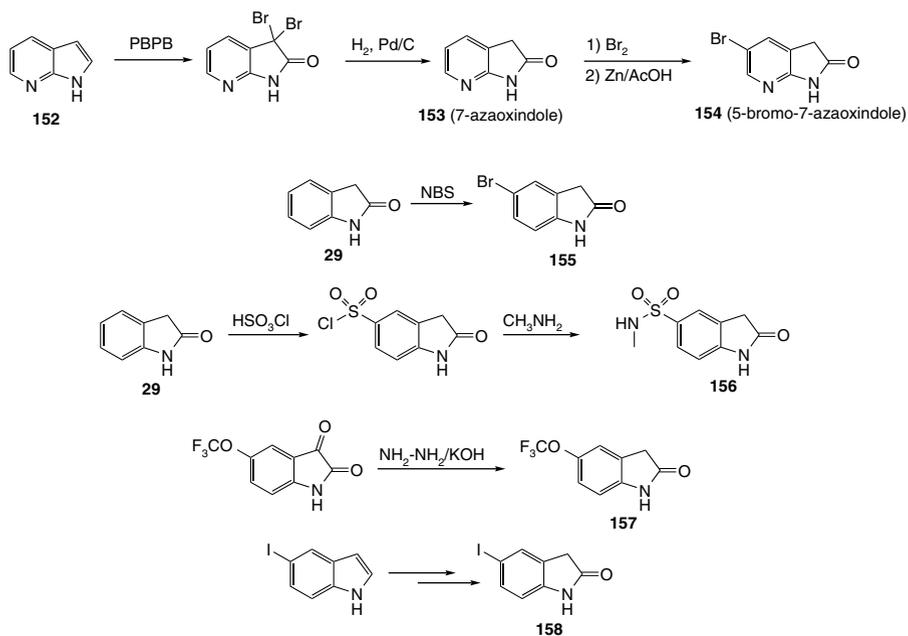
Scheme 47 Preparations of selected intermediates
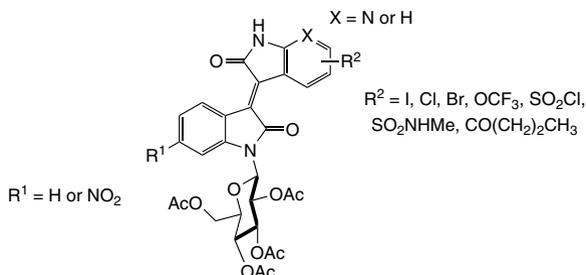
Fig. 29 New assembly of isoindigo and 7′-azaisoindigo derivatives

To fulfill this objective, the corresponding 7-azaoxindoles have been prepared from 7-azaindole (**152**) (*204*) using pyridinium bromide perbromide (PBPB) following by catalytic hydrogenation. 7-Azaoxindole (**153**) has then served as a scaffold for the two-step introduction of bromine at the 5-position to form 5-bromo-7-azaoxindole (**154**). The investigators developed also various substituted oxindoles from simple oxindoles (*e.g.* **155**–**156**), the oxidation of indoles (*e.g.* **157**) or the selective reduction of isatins (*e.g.* **158**). The intermediates were then coupled with unsubstituted or substituted isatins to generate the resultant compound assembly.

Interestingly, the 5′-iodo-7′-azaisoindigo (**159**) prepared has been used lately as a key intermediate for a *Sonogashira* cross-coupling (*205*) in order to introduce various vinyl groups (Scheme 48).
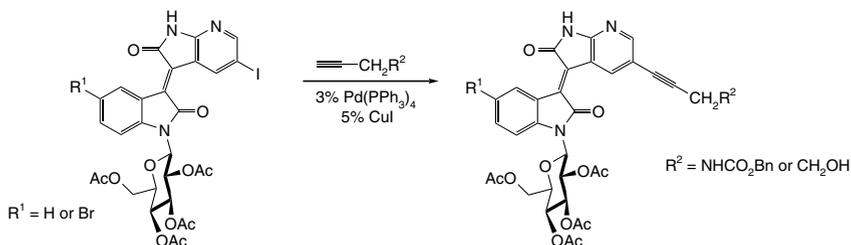
Scheme 48 *Sonogashira* cross-coupling products from 5-iodo-7′-azaisoindigo (**159**)

5.6.2.3 *N*-Substituted Isoindigo

As depicted previously, meisoindigo (**147**) (1-methylisoindigo) possesses an interesting cytotoxic profile, which has led later to the development of a new series of *N*-arylisoindigo derivatives (**160**) (*185*). The targeted compounds were synthesized using a microwave-assisted coupling reaction between a N-substituted isatin and oxindole (**29**) (Scheme 49). The introduction of diverse aryl substituents has been performed by simple nucleophilic substitution between isatin (**27**) and the corresponding halides under microwave irradiation. However, the target compounds exhibited low to moderate cytotoxic potency.

R = CH$_3$; C$_6$H$_5$CH$_2$; C$_6$H$_4$(CH$_2$)$_2$; *p*-CH$_3$C$_6$H$_4$(CH$_2$)$_2$; *p*-CH$_3$OC$_6$H$_4$(CH$_2$)$_2$
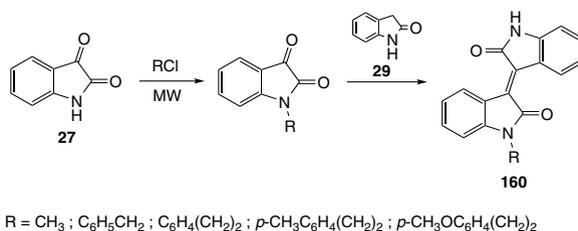
Scheme 49 Synthesis of *N*-arylisoindigos (**160**)

6 Conclusion

"Progress in the Chemistry of Organic Natural Compounds" is an excellent medium to document developments made in the investigation of the indigoid family of natural products. Dyes containing indigoids have been used in societies all around the world to symbolize power (Tyrian purple) and religious seniority (indigo, Tyrian Purple), but also as the color of both working class (indigo) and fashionable clothes (indigo). The commercial and industrial importance of these stimulated early chemists to develop skills to perform the sustainable synthesis of indigoids at low-cost that led to substantial benefits. Approaches to the synthesis of these compounds progressed alongside the concomitant availability of cheap starting materials, first from nitrobenzene, leading eventually biomimetic synthesis involving indoles and isatins. The complexity of the natural production of indigoids interested biochemists, who applied knowledge in their field to assist chemists in understanding their mechanism of formation. Chemists then acquired the ability to synthesize indigo and indirubin using enzymes (*e.g.* plant tissue cultures or cytochrome P450) or by developing chemical catalytic systems based on enzyme functions (*e.g.* cobalt-mediated oxidation).

Although indigoids were related closely to the dye chemistry, studies on traditional Chinese medicine led to the unexpected discovery that indigoids have the ability to alleviate the symptoms of certain types of leukemia, thus projecting interest in this class of natural products and their analogs in the field of medicinal chemistry. The indirubins and bromoindirubins became useful lead compounds in the development of selective protein kinases inhibitors. The availability of simple and versatile synthesis methods allowed the creation of a large assembly of indirubin derivatives (more than 400 compounds), providing an important group of bioactive substances to investigate in various biological activities, such as their effects on stem cells, in parasitology, and their potential use for cancer. Although indigotin is not an active compound, the discovery of akashins and the synthesis of their analogs has enhanced scientific interest in the indigoids. This work involved the development of various derivatives of isoindigo, a compound for a long time considered mainly as by-product that tended to alter the quality of the indigo dye. Moreover, the fascinating electronic structures of indigoids have been applied in material chemistry as new conductors (not covered in this review), and isoindigo-based polymers could

serve as key components for solar panels in the future. In fact, indigoids have a place in many diverse branches of science, including organic chemistry, medicinal chemistry, material and physical chemistry, dye chemistry, biology, and textile science as well as the pictorial arts. There is no doubt that, in the future, members of the colorful family of indigoids will continue to have promising future applications for the benefit of mankind!

References

1. Henshilwood C, d'Errico F, Vanhaeren M, van Niekerk K, Jacobs Z (2004) Middle Stone Age Shell Beads from South Africa. Science 304:404
2. Ferreira ESB, Hulme AN, McNab H, Quye A (2004) The Natural Constituents of Historical Textile Dyes. Chem Soc Rev 33:329
3. Eisner T, Nowicki S, Goetz M, Meinwald J (1980) Red Cochineal Dye (Carminic Acid): Its Role in Nature. Science 208:1039
4. Jurasekova Z, del Puerto E, Bruno G, García-Ramos JV, Sanchez-Cortes S, Domingo C (2010) Extractionless Non-Hydrolysis Surface-Enhanced Raman Spectroscopic Detection of Historical Mordant Dyes on Textile Fibers. J Raman Spectrosc 41:1455
5. Cerrato A, De Santis D, Moresi M (2002) Production of Luteolin Extracts from *Reseda luteola* and Assessment of Their Dyeing Properties. J Sci Food Agric 82:1189
6. Zvezdina SV, Berezin MB, Berezin BD (2010) Natural Dyes Based on Chlorophyll and Protoporphyrin Derivatives. Russ J Coord Chem 36:711
7. Traditional Production of Indigo in France (in French and English). http://www.bleu-de-lectour.com.
8. *Inge Boesken Kanold* Works. http://www.pourpre.inge.free.fr.
9. Karapanagiotis I, de Villemereuil V, Magiatis P, Polychronopoulos P, Vougogiannopoulou K, Skaltsounis AL (2006) Identification of the Coloring Constituents of Four Natural Indigoid Dyes. J Liq Chromatogr Relat Technol 29:1491
10. Surowiec I, Nowik W, Moritz T (2012) Mass Spectrometric Identification of New Minor Indigoids in Shellfish Purple Dye from *Hexaplex trunculus*. Dyes Pigm 94:363
11. Bandaranayake WM (2006) The Nature and Role of Pigments of Marine Invertebrates. Nat Prod Rep 23:223
12. Maskey RP, Grün-Wollny I, Fiebig HH, Laatsch H (2002) Akashins A, B, and C: Novel Chlorinated Indigoglycosides from *Streptomyces* sp. GW 48/1497. Angew Chem Int Ed 41:597
13. Weinstein J, Wyman GM (1956) Spectroscopic Studies on Dyes. I. The Association of Indigo Dyes in the Solid Phase. J Am Chem Soc 78:2387
14. Klessinger M, Lüttke W (1963) Theoretische und Spektroskopische Untersuchungen an Indigo-Farbstoffen: Das Chromophore System der Indigo-Farbstoffe. Tetrahedron 19, Supplement 2:315
15. Jacquemin D, Preat J, Wathelet V, Perpete EA (2006) Substitution and Chemical Environment Effects on the Absorption Spectrum of Indigo. J Chem Phys 124:074104
16. Maugard T, Enaud E, Choisy P, Legoy MD (2001) Identification of an Indigo Precursor from Leaves of *Isatis tinctoria* (Woad). Phytochemistry 58:897
17. Christie RM (2007) Why Is Indigo Blue? Biotech Histochem 82:51

18. Christie RM (2006) Indirubin: The Relationship Between Chemical Structure and Physical Properties, With Particular Reference to Color. In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 103

19. Yao M, Araki M, Senoh H, Yamazaki S-i, Sakai T, Yasuda K (2010) Indigo Dye as a Positive-Electrode Material for Rechargeable Lithium Batteries. Chem Lett 39:950

20. Irimia-Vladu M, Głowacki ED, Troshin PA, Schwabegger G, Leonat L, Susarova DK, Krystal O, Ullah M, Kanbur Y, Bodea MA, Razumov VF, Sitter H, Bauer S, Sariciftci NS (2012) Indigo - A Natural Pigment for High Performance Ambipolar Organic Field Effect Transistors and Circuits. Adv Mater 24:375

21. Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT (1983) Expression of Naphthalene Oxidation Genes in *Escherichia coli* Results in the Biosynthesis of Indigo. Science 222:167

22. Balfour-Paul J (1998) Indigo. British Museum Press, London

23. Giustetto R, Llabres i Xamena FX, Ricchiardi G, Bordiga S, Damin A, Gobetto R, Chierotti MR (2005) Maya Blue: A Computational and Spectroscopic Study. J Phys Chem B 109:19360

24. Ricketts R (2006) *Polygonum tinctorium*: Contemporary Indigo Farming and Processing in Japan. In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 7

25. Zech-Matterne V, Leconte L (2010) New Archaeobotanical Finds of *Isatis tinctoria* L. (Woad) from Iron Age Gaul and a Discussion of The Importance of Woad in Ancient Time. Veg Hist Archaeobot 19:137

26. Schunck E (1900) New Investigations About Indican. Chem News J Ind Sci 82:176

27. Epstein E, Nabors MW, Stowe BB (1967) Origin of Indigo of Woad. Nature 216:547

28. Gilbert KG, Hill DJ, Crespo C, Mas A, Lewis M, Rudolph B, Cooke DT (2000) Qualitative Analysis of Indigo Precursors from Woad by HPLC and HPLC-MS. Phytochem Anal 11:18

29. Oberthür C, Schneider B, Graf H, Hamburger M (2004) The Elusive Indigo Precursors in Woad (*Isatis tinctoria* L.) – Identification of the Major Indigo Precursor, Isatan A, and a Structure Revision of Isatan B. Chem Biodiversity 1:174

30. Maier W, Schumann B, Gröger D (1990) Biosynthesis of Indoxyl Derivatives in *Isatis tinctoria* and *Polygonum tinctorium*. Phytochemistry 29:817

31. Xia Z-Q, Zenk MH (1992) Biosynthesis of Indigo Precursors in Higher Plants. Phytochemistry 31:2695

32. Oberthür C, Graf H, Hamburger M (2004) The Content of Indigo Precursors in *Isatis tinctoria* Leaves, a Comparative Study of Selected Accessions and Post-Harvest Treatments. Phytochemistry 65:3261

33. Garcia-Macias P, John P (2004) Formation of Natural Indigo Derived from Woad (*Isatis tinctoria* L.) in Relation to Product Purity. J Agric Food Chem 52:7891

34. Rousseau X, Vragar Y (2004) Les Indigoteries de Marie-Galante. J Carib Arch 1:47

35. Yvon T (2007) La Production d'Indigo en Guadeloupe au XVIIème et XVIIIème siècle ou l'Archéologie d'une des Premières Industries du Nouveau Monde. http://www.medieval-europe-paris-2007.univ-paris1.fr/T.Yvon.pdf.

36. Laitonjam WS, Wangkheirakpam SD (2011) Comparative Study of the Major Components of the Indigo Dye Obtained from *Strobilanthes flaccidifolius* Nees. and *Indigofera tinctoria* Linn. Int J Plant Phys Biochem 3:108

37. Chanayath N, Lhieochaiphant S, Phutrakul S (2002) Pigment Extraction Techniques from the Leaves of *Indigofera tinctoria* Linn. and *Baphicacanthus cusia* Brem. and Chemical Structure Analysis of Their Major Components. CMJ Journal 1:149

38. Wu E, Komolpis K, Wang H (1999) Chemical Extraction of Indigo from *Indigofera tinctoria* While Attaining Biological Integrity. Biotechnol Tech 13:567

39. Baeyer A, Emmerling A (1870) Reduction des Isatins zu Indigblau. Ber Chem Ges 3:514

40. Baeyer A (1879) Ueber die Einwirkung des Fünffachchlorphosphors auf Isatin und auf Verwandte Substanzen. Ber Chem Ges 12:456

41. Baeyer A (1880) Ueber die Beziehungen der Zimmtsäure zu der Indigogruppe. Ber Chem Ges 13:2254

42. Baeyer A, Drewsen V (1882) Darstellung von Indigblau aus Orthonitrobenzaldehyd. Ber Chem Ges 15:2856
43. Gosteli J (1977) New Indigo Syntheses. Preliminary Communication. Helv Chim Acta 60:1980
44. Cooksey CJ (2007) Indigo: An Annotated Bibliography. Biotech Histochem 82:105
45. Holt SJ, Sadler PW (1958) Studies in Enzyme Cytochemistry. III. Relationships between Solubility, Molecular Association and Structure in Indigoid Dyes. Proc R Soc Lond B Biol 148:495
46. Russell GA, Kaupp G (1969) Reactions of Resonance Stabilized Carbanions. XXXI. Oxidation of Carbanions. 4. Oxidation of Indoxyl to Indigo in Basic Solution. J Am Chem Soc 91:3851
47. Haubrichs R (2006) Natural History and Iconography of Purple Shells. In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 55
48. Discovery of Tyrian Purple by *Hercules*. http://www.en.wikipedia.org/wiki/File:Peter_Paul_ Rubens_-_La_d%C3%A9couverte_de_la_pourpre.JPG.
49. Sotiropoulou S (2004) La Pourpre dans l'Art Cycladique: Identification du Pigment dans les Peintures Murales d'Akrotiri (Théra, Grèce). Preist Alp 40:167
50. Sotiropoulou S, Karapanagiotis I (2006) Conchylian Purple Investigation in Prehistoric Wall Paintings of the Aegean Area. In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 71
51. Karapanagiotis I (2006) Identification of Indigoid Natural Dyestuffs Used in Art Objects by HPLC Coupled to APCI-MS. Am Lab 38:38
52. McGovern PE, Michel RH (1985) Royal Purple Dye: Tracing Chemical Origins of the Industry. Anal Chem 57:1514A
53. Koren ZC (2008) Archaeo-Chemical Analysis of Royal Purple on a Darius I Stone Jar. Microchim Acta 162:381
54. McGovern PE, Michel RH (1990) Royal Purple Dye: The Chemical Reconstruction of the Ancient Mediterranean Industry. Acc Chem Res 23:152
55. Haubrichs R (2004) L'Etude de la Pourpre: Histoire d'une Couleur, Chimie et Expérimentation. Preist Alp 40:3
56. Cole W (1685) A Letter from Mr *William Cole* of Bristol, to the Phil. Society of Oxford; Containing His Observations on the Purple Fish. Phil Trans 15:1278
57. Cooksey CJ (2001) Tyrian Purple: 6,6′-Dibromoindigo and Related Compounds. Molecules 6:736
58. Terada T (2008) Paper 138: Sea Snails in Contemporary Japanese Embroidery. In: American Textiles Society Symposium Proceedings, Textile Society of America 11th Biennial Symposium, Honolulu, Hawaii, USA, 4–7 September 2008, p 1
59. Michel RH, McGovern PE (1987) The Chemical Processing of Royal Purple Dye: Ancient Descriptions as Elucidated by Modern Science. Archeomaterials 1:135
60. Lopez Chavez FJ, Chavez PR, Oyama K (2009) Brominated Precursors of Tyrian purple (C.I. Natural Violet 1) from *Plicopurpura pansa*, *Plicopurpura columellaris* and *Plicopurpura patula*. Dyes Pigm 83:7
61. Michel RH, Lazar J, McGovern PE (1992) The Chemical Composition of the Indigoid Dyes Derived from The Hypobranchial Glandular Secretions of *Murex* Molluscs. J Soc Dyers Colour 108:145
62. Meijer L, Skaltsounis A-L, Magiatis P, Polychronopoulos P, Knockaert M, Leost M, Ryan XP, Vonica CA, Brivanlou A, Dajani R, Crovace C, Tarricone C, Musacchio A, Roe SM, Pearl L, Greengard P (2003) GSK-3 Selective Inhibitors Derived from Tyrian Purple Indirubins. Chem Biol 10:1255
63. Clark RJH, Cooksey CJ (1997) Bromoindirubins: The Synthesis and Properties of Minor Components of Tyrian Purple and The Composition of the Colorant from *Nucella lapillus*. J Soc Dyers Colour 113:316
64. Koren ZC (1994) HPLC Analysis of the Natural Scale Insect, Madder and Indigoid Dyes. J Soc Dyers Colour 110:273
65. Baker JT, Sutherland MD (1968) Pigments of Marine Animals: VIII. Precursors of 6,6′-Dibromoindigotin (Tyrian Purple) from the Mollusc *Dicathais orbita* Gmelin. Tetrahedron Lett 9:43

66. Baker JT, Duke CC (1973) Isolation from Marine Mollusks Hypobranchial Glands of 6-Bromo-2,2-bis(Methylthio)indolin-3-one and 6-Bromo-2-(methylthio)indoleninone as Alternative Precursors to Tyrian Purple. Tetrahedron Lett 27:2481

67. Fouquet H, Bielig H-J (1971) Biological Precursors and Genesis of Tyrian Purple. Angew Chem Int Ed 10:816

68. Cooksey CJ (2006) Marine Indirubins. In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 23

69. Baker JT, Duke CC (1976) Isolation of Choline and Choline Ester Salts of Tyrindoxyl Sulfate from the Marine Molluscs, *Dicathais orbita* and *Mancinella keineri*. Tetrahedron Lett 15:1233

70. Westley C, Benkendorff K (2008) Sex-Specific Tyrian Purple Genesis: Precursor and Pigment Distribution in the Reproductive System of the Marine Mollusc *Dicathais orbita*. J Chem Ecol 34:44

71. Westley CB, Vine KL, Benkendorff K (2006) A Proposed Functional Role for Indole Derivatives in Reproduction and Defence of the Muricidae (Neogastropoda: Mollusca) In: Indirubin, The Red Shade of Indigo. Life in Progress Editions. Station Biologique, Roscoff, France, p 31

72. Jimenez EC, Craig AG, Watkins M, Hillyard DR, Gray WR, Gulyas J, Rivier JE, Cruz LJ, Olivera BM (1997) Bromocontryphan: A Post-Translational Bromination of Tryptophan. Biochemistry 36:989

73. Ten Brink HB, Schoemaker HE, Wever R (2001) Sulfoxidation Mechanism of Vanadium Bromoperoxidase from *Ascophyllum nodosum*. Eur J Biochem 268:132

74. Benkendorff K, Bremner J, Davis A (2000) Tyrian Purple Precursors in the Egg Masses of the Australian Muricid, *Dicathais orbita*: A Possible Defensive Role. J Chem Ecol 26:1037

75. Christophersen C, Wätjen F, Buchardt O, Anthoni U (1978) A Revised Structure of Tyriverdin: The Precursor of Tyrian Purple. Tetrahedron 34:2779

76. Fujise Y, Miwa K, Ito S (1980) Structure of Tyriverdin, The Immediate Precursor of Tyrian Purple. Chem Lett 9:631

77. Michel RH, McGovern PE (1990) The Chemical Processing of Royal Purple Dye: Ancient Descriptions as Elucidated by Modern Science, Part II. Archeomaterials 4:97

78. Sachs F, Kempf R (1903) Ueber *p*-Halogen-*o*-Nitrobenzaldehyde. Ber Chem Ges 36:3299

79. Sachs F, Sichel E (1904) Ueber *p*-Substituirte-*o*-Nitrobenzaldehyde. Ber Chem Ges 37:1861

80. Friedländer P (1909) Über den Farbstoff des Antiken Purpurs aus *Murex brandaris*. Ber Chem Ges 42:765

81. Ettinger L, Friedländer P (1912) Über *N*-Methyl-Derivate des Indigblaus. Ber Chem Ges 45:2074

82. Grandmougin E, Seyder P (1914) Über Indigo. V: Über Halogenierte Indigo und Derivate. Ber Chem Ges 47:2365

83. Barber HJ, Stickings CE (1945) 41. Phenanthrene-3,6-diamidine. J Chem Soc167

84. Voß G, Gerlach H (1989) Regioselektiver Brom/Lithium-Austausch bei 2,5-Dibrom-1-Nitrobenzol. – Eine Einfache Synthese von 4-Brom-2-Nitrobenzaldehyd und 6,6′-Dibromindigo. Chem Ber 122:1199

85. Majima R, Kotake M (1930) Synthetische Versuche in der Indol-Gruppe, VII: Über Nitrieren und Bromieren des β-Indol-Carbonsäure-Esters und Eine Neue Synthese des Farbstoffs des Antiken Purpurs. Ber Chem Ges 63:2237

86. Wolk JL, Frimer AA (2010) A Simple, Safe and Efficient Synthesis of Tyrian Purple (6,6′-Dibromoindigo). Molecules 15:5561

87. Tanoue Y, Terada A, Sakata K, Hashimoto M, Morishita S-I, Hamada M, Kai N (2001) A Facile Synthesis of Tyrian Purple Based on a Biosynthetic Pathway. Fish Sci 67:726

88. Arnold R, Nutter W, Stepp W (1959) Indoxyl Acetate from Indole. J Org Chem 24:117

89. Hoessel R, Leclerc S, Endicott JA, Nobel MEM, Lawrie A, Tunnah P, Leost M, Damiens E, Marie D, Marko D, Niederberger E, Tang W, Eisenbrand G, Meijer L (1999) Indirubin, the Active Constituent of a Chinese Antileukaemia Medicine, Inhibits Cyclin-Dependent Kinases. Nat Cell Biol 1:60

90. Damiens E, Baratte B, Marie D, Eisenbrand G, Meijer L (2001) Anti-Mitotic Properties of Indirubin-3′-monoxime, a CDK/GSK-3 Inhibitor: Induction of Endoreplication Following Prophase Arrest. Oncogene 20:3786

91. Marko D, Schatzle S, Friedel A, Genzlinger A, Zankl H, Meijer L, Eisenbrand G (2001) Inhibition of Cyclin-Dependent Kinase 1 (CDK1) by Indirubin Derivatives in Human Tumour Cells. Br J Cancer 84:283

92. Xiao Z, Hao Y, Liu B, Qian L (2002) Indirubin and Meisoindigo in the Treatment of Chronic Myelogenous Leukemia in China. Leuk Lymphoma 43:1763

93. Nam S, Buettner R, Turkson J, Kim D, Cheng JQ, Muehlbeyer S, Hippe F, Vatter S, Merz K-H, Eisenbrand G, Jove R (2005) Indirubin Derivatives Inhibit STAT3 Signaling and Induce Apoptosis in Human Cancer Cells. Proc Natl Acad Sci USA 102:5998

94. Senderowicz AM (2000) Small Molecule Modulators of Cyclin-Dependent Kinases for Cancer Therapy. Oncogene 19:6600

95. Nabel EG (2002) CDKs and CKIs: Molecular Targets for Tissue Remodelling. Nat Rev Drug Discov 1:587

96. Malumbres M, Barbacid M (2009) Cell Cycle, CDKs and Cancer: A Changing Paradigm. Nat Rev Cancer 9:153

97. Leclerc S, Garnier M, Hoessel R, Marko D, Bibb JA, Snyder GL, Greengard P, Biernat J, Wu Y-Z, Mandelkow E-M, Eisenbrand G, Meijer L (2001) Indirubins Inhibit Glycogen Synthase Kinase-3β and CDK5/P25, Two Protein Kinases Involved in Abnormal Tau Phosphorylation in Alzheimer's Disease: A Property Common to Most Cyclin-Dependent Kinase Inhibitors? J Biol Chem 276:251

98. Davies TG, Tunnah P, Meijer L, Marko D, Eisenbrand G, Endicott JA, Noble MEM (2001) Inhibitor Binding to Active and Inactive CDK2: The Crystal Structure of CDK2-Cyclin A/ Indirubin-5-sulfonate. Structure 9:389

99. Vougogiannopoulou K, Skaltsounis AL (2012) From Tyrian Purple to Kinase Modulators: Naturally Halogenated Indirubins and Synthetic Analogues. Planta Med 78:1515

100. Rayasam GV, Tulasi VK, Sodhi R, Davis JA, Ray A (2009) Glycogen Synthase Kinase-3: More Than a Namesake. Br J Pharmacol 156:885

101. Srivastava A, Pandey S (1998) Potential Mechanism(s) Involved in the Regulation of Glycogen Synthesis by Insulin. Mol Cell Biochem 182:135

102. Ross SA, Gulve EA, Wang M (2004) Chemistry and Biochemistry of Type 2 Diabetes. Chem Rev 104:1255

103. Martinez A, Castro A, Dorronsoro I, Alonso M (2002) Glycogen Synthase Kinase-3 (GSK-3) Inhibitors as New Promising Drugs for Diabetes, Neurodegeneration, Cancer, and Inflammation. Med Res Rev 22:373

104. Medina M, Avila J (2010) Glycogen Synthase Kinase-3 (GSK-3) Inhibitors for the Treatment of Alzheimer's Disease. Curr Pharm Des 16:2790

105. Mondragón-Rodríguez S, Perry G, Zhu X, Moreira PI, Williams S (2012) Glycogen Synthase Kinase 3: A Point of Integration in *Alzheimer's* Disease and a Therapeutic Target? Int J *Alzheimer*'s Dis 2012:4

106. Picchini AM, Manji HK, Gould TD (2004) GSK-3 and Neurotrophic Signaling: Novel Targets Underlying The Pathophysiology and Treatment of Mood Disorders? Drug Discov Today Dis Mech 1:419

107. Freyberg Z, Ferrando SJ, Javitch JA (2010) Roles of the Akt/GSK-3 and Wnt Signaling Pathways in Schizophrenia and Antipsychotic Drug Action. Am J Psychiat 167:388

108. Takenaka K, Kise Y, Miki H (2007) GSK-3beta Positively Regulates Hedgehog Signaling Through Sufu in Mammalian Cells. Biochem Biophys Res Commun 353:501

109. Toledo EM, Colombres M, Inestrosa NC (2008) Wnt Signaling in Neuroprotection and Stem Cell Differentiation. Prog Neurobiol 86:281

110. Polakis P (2012) Drugging Wnt Signalling in Cancer. EMBO J 31:2737

111. Fearon ER, Spence JR (2012) Cancer Biology: A New RING to Wnt Signaling. Curr Biol 22:R849

112. Coni S, Infante P, Gulino A (2013) Control of Stem Cells and Cancer Stem Cells by Hedgehog Signaling: Pharmacologic Clues from Pathway Dissection. Biochem Pharmacol 85:623

113. McMillan R, Matsui W (2012) Molecular Pathways: The Hedgehog Signaling Pathway in Cancer. Clin Cancer Res 18:4883

114. Beurel E, Jope RS (2008) Differential Regulation of STAT Family Members by Glycogen Synthase Kinase-3. J Biol Chem 283:21934

115. Polychronopoulos P, Magiatis P, Skaltsounis A-L, Myrianthopoulos V, Mikros E, Tarricone A, Musacchio A, Roe SM, Pearl L, Leost M, Greengard P, Meijer L (2004) Structural Basis for the Synthesis of Indirubins as Potent and Selective Inhibitors of Glycogen Synthase Kinase-3 and Cyclin-Dependent Kinases. J Med Chem 47:935

116. Martin L, Magnaudeix A, Esclaire F, Yardin C, Terro F (2009) Inhibition of Glycogen Synthase Kinase-3β Downregulated Total Tau Proteins in Cultured Neurons and Its Reversal by the Blockade of Protein Phosphatase-2A. Brain Res 1252:66

117. Knockaert M, Blondel M, Bach S, Leost M, Elbi C, Hager GL, Nagy SR, Han D, Denison M, Ffrench M, Ryan XP, Magiatis P, Polychronopoulos P, Greengard P, Skaltsounis L, Meijer L (2004) Independent Actions on Cyclin-Dependent Kinases and Aryl Hydrocarbon Receptor Mediate the Antiproliferative Effects of Indirubins. Oncogene 23:4400

118. Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH (2003) Molecular Signature of Human Embryonic Stem Cells and Its Comparison with the Mouse. Dev Biol 260:404

119. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH (2004) Maintenance of Pluripotency in Human and Mouse Embryonic Stem Cells Through Activation of Wnt Signaling by a Pharmacological GSK-3-Specific Inhibitor. Nat Med 10:55

120. Sato H, Amagai K, Shimizukawa R, Tamai Y (2009) Stable Generation of Serum- and Feeder-Free Embryonic Stem Cell-Derived Mice with Full Germline-Competency by Using a GSK-3 Specific Inhibitor. Genesis 47:414

121. Hammarton TC, Mottram JC, Doerig C (2003) The Cell Cycle of Parasitic Protozoa: Potential for Chemotherapeutic Exploitation. Prog Cell Cycle Res 5:91

122. Naula C, Parsons M, Mottram JC (2005) Protein Kinases as Drug Targets in Trypanosomes and Leishmania. Biochim Biophys Acta 1754:151

123. WHO (2010) Control of Leishmaniases. WHO Technical Report Series, vol 949, Geneva

124. Ojo KK, Arakaki TL, Napuli AJ, Inampudi KK, Keyloun KR, Zhang L, Hol WG, Verlinde CL, Merritt EA, Van Voorhis WC (2011) Structure Determination of Glycogen Synthase Kinase-3 from *Leishmania major* and Comparative Inhibitor Structure-Activity Relationships with *Trypanosoma brucei* GSK-3. Mol Biochem Parasitol 176:98

125. Ojo KK, Gillespie JR, Riechers AJ, Napuli AJ, Verlinde CL, Buckner FS, Gelb MH, Domostoj MM, Wells SJ, Scheer A, Wells TN, Van Voorhis WC (2008) Glycogen Synthase Kinase-3 Is A Potential Drug Target for African Trypanosomiasis Therapy. Antimicrob Agents Chemother 52:3710

126. Hassan P, Fergusson D, Grant KM, Mottram JC (2001) The CRK3 Protein Kinase is Essential for Cell Cycle Progression of *Leishmania mexicana*. Mol Biochem Parasitol 113:189

127. Grant KM, Dunion MH, Yardley V, Skaltsounis AL, Marko D, Eisenbrand G, Croft SL, Meijer L, Mottram JC (2004) Inhibitors of *Leishmania mexicana* CRK3 Cyclin-Dependent Kinase: Chemical Library Screen and Antileishmanial Activity. Antimicrob Agents Chemother 48:3033

128. Xingi E, Smirlis D, Myrianthopoulos V, Magiatis P, Grant KM, Meijer L, Mikros E, Skaltsounis A-L, Soteriadou K (2009) 6-Br-5-Methylindirubin-3′-oxime (5-Me-6-BIO) Targeting the Leishmanial Glycogen Synthase Kinase-3 (GSK-3) Short Form Affects Cell-Cycle Progression and Induces Apoptosis-Like Death: Exploitation of GSK-3 for Treating Leishmaniasis. Int J Parasitol 39:1289

129. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The Protein Kinase Complement of the Human Genome. Science 298:1912

130. Levitzki A, Klein S (2010) Signal Transduction Therapy of Cancer. Mol Aspects Med 31:287

131. Ribas J, Bettayeb K, Ferandin Y, Knockaert M, Garrofe-Ochoa X, Totzke F, Schachtele C, Mester J, Polychronopoulos P, Magiatis P, Skaltsounis AL, Boix J, Meijer L (2006) 7-Bromoindirubin-3′-oxime Induces Caspase-Independent Cell Death. Oncogene 25:6304

132. Liu L, Nam S, Tian Y, Yang F, Wu J, Wang Y, Scuto A, Polychronopoulos P, Magiatis P, Skaltsounis L, Jove R (2011) 6-Bromoindirubin-3′-oxime Inhibits JAK/STAT3 Signaling and Induces Apoptosis of Human Melanoma Cells. Cancer Res 71:3972

133. Nicolaou KA, Liapis V, Evdokiou A, Constantinou C, Magiatis P, Skaltsounis AL, Koumas L, Costeas PA, Constantinou AI (2012) Induction of Discrete Apoptotic Pathways by Bromo-Substituted Indirubin Derivatives in Invasive Breast Cancer Cells. Biochem Biophys Res Commun 425:76

134. Vougogiannopoulou K, Ferandin Y, Bettayeb K, Myrianthopoulos V, Lozach O, Fan Y, Johnson CH, Magiatis P, Skaltsounis A-L, Mikros E, Meijer L (2008) Soluble 3′,6-Substituted Indirubins with Enhanced Selectivity toward Glycogen Synthase Kinase-3 Alter Circadian Period. J Med Chem 51:6421

135. Martin L, Magnaudeix A, Wilson CM, Yardin C, Terro F (2011) The New Indirubin Derivative Inhibitors of Glycogen Synthase Kinase-3, 6-BIDECO and 6-BIMYEO, Prevent Tau Phosphorylation and Apoptosis Induced by the Inhibition of Protein Phosphatase-2A by Okadaic Acid in Cultured Neurons. J Neurosci Res 89:1802

136. Schunck E (1879) LV.-On Indigo-Purpurin and Indirubin. J Chem Soc Trans 35:528

137. Baeyer A (1878) Synthese des Indigblaus. Ber Chem Ges 11:1296

138. Baeyer A (1883) Ueber die Verbindungen der Indigogruppe. Ber Chem Ges 16:2188

139. Wahl A, Bagard P (1910) New Synthesis of Indirubin. Bull Soc Chim Fr 5:1043

140. Katritzky AR, Fan W-Q, Koziol AE, Palenik GJ (1989) 2-Chloro-3*H*-indol-3-one and Its Reactions with Nucleophiles. J Heterocycl Chem 26:821

141. Wahl A, Bagard P (1914) Syntheses in the Indigoid Group. II. New Syntheses of Scarlot Thioindigo and Indirubin. Bull Soc Chim Fr 15:pp. 336

142. Martinet J (1919) Sur les Indirubines. C R Hebd Seances Acad Sci 169:183

143. Albert A (1915) Thioindol-Derivate. Eine Synthese des Indirubins. Schwefelhaltige Indigoide Farbstoffe I. Ber Chem Ges 48:474

144. Ainley AD, Robinson R (1934) The Epindoline Group. Part I. Trial of Various Methods for the Synthesis of Epindolidiones. J Chem Soc:1508

145. Giovannini E, Lorenz T (1957) Reduktionen mit LiAlH$_4$ in der Isatin-Reihe. 2. Mitteilung. Einwirkung von LiAlH$_4$ auf Indol- und Isatin-Derivate. Helv Chim Acta 40:2287

146. Thomas F, Bloxam WP, Perkin AG (1909) XCII.-Indican. Part III. J Chem Soc Trans 95:824

147. Perkin AG (1909) XCIII.-Indoxylic Acid. J Chem Soc Trans 95:847

148. Perkin AG, Bloxam WP (1907) XXX.-Some Constituents of Natural Indigo. Part I. J Chem Soc Trans 91:279

149. Martinet J, Dornier O (1920) Sur les Azoïques de l'Indoxyle. C R Hebd Seances Acad Sci 170:592

150. Vorländer D, von Pfeiffer J (1919) Über Diacetyl-Indigo. Ber Chem Ges 52:325

151. Spencer G (1931) The Hydrolysis of Actylindoxyl Acids and the Acetylindoxyls. J Soc Chem Ind 50:63T

152. Shim J-Y, Chang Y-J, Kim S-U (1998) Indigo and Indirubin Derivatives from Indoles in *Polygonum tinctorium* Tissue Cultures. Biotechnol Lett 20:1139

153. Guengerich FP, Sorrells JL, Schmitt S, Krauser JA, Aryal P, Meijer L (2004) Generation of New Protein Kinase Inhibitors Utilizing Cytochrome P450 Mutant Enzymes for Indigoid Synthesis. J Med Chem 47:3236

154. Das PJ, Baruah A (2011) Oxidative Degradation of Indole Using Mixed Ligand Co(II) Complex and Hydrogen Peroxide. Selective Formation of Indole Dimers. Chem Biol Interface 1:355

155. Shivankar VS, Thakkar NV (2005) Decomposition of Hydrogen Peroxide in Presence of Mixed Ligand Cobalt(II) and Nickel(II) Complexes as Catalysts. J Sci Ind Res 64:496

156. Kim S-A, Kim Y-C, Kim S-W, Lee S-H, Min J-J, Ahn S-G, Yoon J-H (2007) Antitumor Activity of Novel Indirubin Derivatives in Rat Tumor Model. Clin Cancer Res 13:253

157. Moon MJ, Lee SK, Lee J-W, Song WK, Kim SW, Kim JI, Cho C, Choi SJ, Kim Y-C (2006) Synthesis and Structure-Activity Relationships of Novel Indirubin Derivatives as Potent Anti-Proliferative Agents with CDK2 Inhibitory Activities. Bioorg Med Chem 14:237

158. Beauchard A, Ferandin Y, Frère S, Lozach O, Blairvacq M, Meijer L, Thiéry V, Besson T (2006) Synthesis of Novel 5-Substituted Indirubins as Protein Kinases Inhibitors. Bioorg Med Chem 14:6434

159. Ferandin Y, Bettayeb K, Kritsanida M, Lozach O, Polychronopoulos P, Magiatis P, Skaltsounis A-L, Meijer L (2006) 3′-Substituted 7-Halogenoindirubins, a New Class of Cell Death Inducing Agents. J Med Chem 49:4638

160. Lee J-W, Moon MJ, Min H-Y, Chung H-J, Park E-J, Park HJ, Hong J-Y, Kim Y-C, Lee SK (2005) Induction of Apoptosis by a Novel Indirubin-5-nitro-3′-monoxime, a CDK Inhibitor, in Human Lung Cancer Cells. Bioorg Med Chem Lett 15:3948

161. Wang ZH, Dong Y, Wang T, Shang MH, Hua WY, Yao QZ (2010) Synthesis and CDK2 Kinase Inhibitory Activity of 7/7′-Azaindirubin Derivatives. Chin Chem Lett 21:297

162. Kritsanida M, Magiatis P, Skaltsounis A-L, Peng Y, Li P, Wennogle LP (2009) Synthesis and Antiproliferative Activity of 7-Azaindirubin-3′-oxime, a 7-Aza Isostere of the Natural Indirubin Pharmacophore. J Nat Prod 72:2199

163. Myrianthopoulos V, Kritsanida M, Gaboriaud-Kolar N, Magiatis P, Ferandin Y, Durieu E, Lozach O, Cappel D, Soundararajan M, Filippakopoulos P, Sherman W, Knapp S, Meijer L, Mikros E, Skaltsounis A-L (2013) Novel Inverse Binding Mode of Indirubin Derivatives Yields Improved Selectivity for DYRK Kinases. ACS Med Chem Lett 4:22

164. Choi S-J, Lee J-E, Jeong S-Y, Im I, Lee S-D, Lee E-J, Lee SK, Kwon S-M, Ahn S-G, Yoon J-H, Han S-Y, Kim J-I, Kim Y-C (2010) 5,5′-Substituted Indirubin-3′-oxime Derivatives as Potent Cyclin-Dependent Kinase Inhibitors with Anticancer Activity. J Med Chem 53:3696

165. Nam S, Scuto A, Yang F, Chen W, Park S, Yoo H-S, Konig H, Bhatia R, Cheng X, Merz K-H, Eisenbrand G, Jove R (2012) Indirubin Derivatives Induce Apoptosis of Chronic Myelogenous Leukemia Cells Involving Inhibition of Stat5 Signaling. Mol Oncol 6:276

166. Sen B, Saigal B, Parikh N, Gallick G, Johnson FM (2009) Sustained Src Inhibition Results in Signal Transducer and Activator of Transcription 3 (STAT3) Activation and Cancer Cell Survival *via* Altered Janus-Activated Kinase, STAT3 Binding. Cancer Res 69:1958

167. Johnson FM, Saigal B, Tran H, Donato NJ (2007) Abrogation of Signal Transducer and Activator of Transcription 3 Reactivation after Src Kinase Inhibition Results in Synergistic Antitumor Effects. Clin Cancer Res 13:4233

168. Nam S, Wen W, Schroeder A, Herrmann A, Yu H, Cheng X, Merz KH, Eisenbrand G, Li H, Yuan YC, Jove R (2013) Dual Inhibition of Janus and Src Family Kinases by Novel Indirubin Derivative Blocks Constitutively-Activated STAT3 Signaling Associated with Apoptosis of Human Pancreatic Cancer Cells. Mol Oncol 7:369

169. Myrianthopoulos V, Magiatis P, Ferandin Y, Skaltsounis A-L, Meijer L, Mikros E (2007) An Integrated Computational Approach to the Phenomenon of Potent and Selective Inhibition of Aurora Kinases B and C by a Series of 7-Substituted Indirubins. J Med Chem 50:4027

170. Choi SJ, Moon MJ, Lee SD, Choi S-U, Han S-Y, Kim Y-C (2010) Indirubin Derivatives as Potent FLT3 Inhibitors with Anti-Proliferative Activity of Acute Myeloid Leukemic Cells. Bioorg Med Chem Lett 20:2033

171. Han S-Y, Ahn JH, Shin CY, Choi S-U (2010) Effects of Indirubin Derivatives on the FLT3 Activity and Growth of Acute Myeloid Leukemia Cell Lines. Drug Dev Res 71:221

172. Pemmaraju N, Kantarjian H, Ravandi F, Cortes J (2011) FLT3 Inhibitors in the Treatment of Acute Myeloid Leukemia. Cancer 117:3293

173. Chan P (2011) Differential Signaling of FLT3 Activating Mutations in Acute Myeloid Leukemia: A Working Model. Protein Cell 2:108

174. Hein M, Michalik D, Langer P (2005) First Synthesis of *N*-Indigoglycosides - Analogues of Cancerostatic Akashines. Synthesis 2005:3531

175. Hein M, Phuong NTB, Michalik D, Görls H, Lalk M, Langer P (2006) Synthesis of the First Deprotected Indigo *N*-Glycosides (Blue Sugars) by Reductive Glycosylation of Dehydroindigo. Tetrahedron Lett 47:5741

176. Libnow S, Hein M, Michalik D, Langer P (2006) First Synthesis of Indirubin *N*-Glycosides (Red Sugars). Tetrahedron Lett 47:6907

177. Libnow S, Methling K, Hein M, Michalik D, Harms M, Wende K, Flemming A, Köckerling M, Reinke H, Bednarski PJ, Lalk M, Langer P (2008) Synthesis of Indirubin-*N*′-Glycosides and Their Anti-Proliferative Activity Against Human Cancer Cell Lines. Bioorg Med Chem 16:5570

178. Libnow S, Hein M, Langer P (2009) The First N-Glycosylated Indoxyls and Their Application to the Synthesis of Indirubin-*N*-Glycosides (Purple Sugars). Synlett 2009:221

179. Eisenbrand G (2000) Preparation of Indigoid Bisindole Indirubin Derivatives for Pharmaceutical Use in the Treatment of Solid Cancers. WO2000061555A1

180. Eisenbrand G, Hippe F (2003) Preparation of Hydroxylated Indirubin Glucuronides as Potential Antitumor Agents. WO2003070703A2

181. Wahl A, Bagard P (1909) Sur un Nouvel Isomère de l'Indigo. C R Hebd Seances Acad Sci 148:716

182. Wahl A, Bagard P (1914) Synthèse dans le Groupe des Indigoïdes. Dérivés de l'Iso-Indigotine et de l'Indirubine. Bull Soc Chim Fr 15:329

183. Papageorgiou C, Borer X (1988) Acid-Catalyzed Rearrangements for a Diastereoselective Entry into a New Fused Hexacyclic Heterocycle: (5*RS*,7a*RS*,12*RS*,14a*RS*)-4,5,7,7a,11,12,14,14a-Octahydro-5,12-dimethyldiindo[1,7-*bc*:1′,7′-*gh*][2,6]naphthyridine. Helv Chim Acta 71:1079

184. Menozzi C, Dalko PI, Cossy J (2007) Studies on Enol Carbonate Chemistry: Stereoselective Construction of Vicinal Quaternary Benzylic Centers in the Bis-Oxindole Series. Heterocycles 72:199

185. Wee XK, Yeo WK, Go M-L, Zhang B, Tan VBC, Lim KM, Tay TE (2009) Synthesis and Evaluation of Functionalized Isoindigos as Antiproliferative Agents. Bioorg Med Chem 17:7562

186. Zhang W, Go M-L (2009) Functionalized 3-Benzylidene-indolin-2-ones: Inducers of NAD(P) H-Quinone Oxidoreductase 1 (NQO1) with Antiproliferative Activity. Bioorg Med Chem 17:2077

187. Kloeck C, Jin X, Choi K, Khosla C, Madrid PB, Spencer A, Raimundo BC, Boardman P, Lanza G, Griffin JH (2011) Acylideneoxoindoles: A New Class of Reversible Inhibitors of Human Transglutaminase 2. Bioorg Med Chem Lett 21:2692

188. Staudinger H, Goldstein J (1916) Aliphatische Diazoverbindungen. 6. Mitteilung: Diphenyl-Diazomethan-Derivate. Ber Chem Ges 49:1923

189. Jackson AH (1965) Synthesis of Oxindole. Chem Ind:1652

190. Voigt E, Meier H (1975) Über die Valenzisomerie von Diazoverbindungen und Diazirinen. Chem Ber 108:3326

191. Lathourakis GE, Litinas KE (1996) Synthesis and Study of 3-(Triphenylphosphoranylidene)-2,3-dihydro-1*H*-indol-2-one. J Chem Soc, Perkin Trans I:491

192. El-Kateb AA, Hennawy IT, Shabana R, Osman FH (1984) Thiation Reactions. III. Thiation of Certain Dicarbonyl Coumpounds. Phosphorus Sulfur 20:329

193. Bergman J, Romero I (2010) Synthesis of Alpha-oxo-sulfines in the Indole Series. J Heterocycl Chem 47:1215

194. Martiis FD (1957) Formazione di Isoindaco da Ossindolo. Ann Chim 47:1238

195. Chovin P (1944) Recherches sur les Lactones Colorées. Isooxyindigo et Benzonaphtyrones. Bull Soc Chim Fr 11:82

196. Laurent A, Erdmann OL (1842) LIII. Untersuchungen Uber den Indigo. J Prakt Chem 25:430

197. Wahl A, Hansen W (1924) Sur la Constitution de l'Isatane et de l'Isatyde. C R Hebd Seances Acad Sci 178:393

198. Wahl A, Hansen W (1924) Transformation et Constitution de la Disulfisatyde. C R Hebd Seances Acad Sci 178:214

199. Teuber H-J, Hohn J (1982) Reactions with Nitrosodisulfonate, 40. Oxidation of Keto-Enols. Chem Ber 115:90

200. Wang L, Liu X, Chen R (2003) Derivatives of Isoindigo, Indigo and Indirubin and Use in Treating Cancer. WO2003051900

201. Sassatelli M, Saab E, Anizon F, Prudhomme M, Moreau P (2004) Synthesis of Glycosyl-Isoindigo Derivatives. Tetrahedron Lett 45:4827

202. Sassatelli M, Bouchikhi F, Messaoudi S, Anizon F, Debiton E, Barthomeuf C, Prudhomme M, Moreau P (2006) Synthesis and Antiproliferative Activities of Diversely Substituted Glycosyl-Isoindigo Derivatives. Eur J Med Chem 41:88
203. Sassatelli M, Bouchikhi F, Aboab B, Anizon F, Fabbro D, Prudhomme M, Moreau P (2007) *In Vitro* Antiproliferative Activities and Kinase Inhibitory Potencies of Glycosyl-Isoindigo Derivatives. Anti-Cancer Drugs 18:1069
204. Bouchikhi F, Anizon F, Moreau P (2008) Synthesis and Antiproliferative Activities of Isoindigo and Azaisoindigo Derivatives. Eur J Med Chem 43:755
205. Bouchikhi F, Anizon F, Moreau P (2009) Synthesis, Kinase Inhibitory Potencies and *in Vitro* Antiproliferative Activity of Isoindigo and 7′-Azaisoindigo Derivatives Substituted by *Sonogashira* Cross-Coupling. Eur J Med Chem 44:2705

Bioactive Heterocyclic Natural Products from Actinomycetes Having Effects on Cancer-Related Signaling Pathways

Masami Ishibashi

Contents

M. Ishibashi (✉)
Graduate School of Pharmaceutical Sciences, Chiba University,
1-8-1 Inohana, Chuo-ku, Chiba 260-8675, Japan
e-mail: mish@chiba-u.jp

1   Introduction

During studies on the search for bioactive natural products from various natural sources, our group has recently become particularly interested in a screening program targeting signaling molecules related to cancer-related biological pathways such as TRAIL, Wnt, and hedgehog signaling (*1*). For these screening studies, natural product extracts have been investigated, including previously unexplored myxomycetes (*2*, *3*), marine nudibranchs (*4*), in addition to medicinal plants collected from south and southeast Asian countries such as Thailand (*5*) and Bangladesh (*6*). Recently, bioactive metabolites of actinomycete strains have been investigated that were collected in several locations mainly in the Chiba, Japan area. Described in this contribution are the isolation and structure elucidation of principally heterocyclic aromatic compounds, along with their biological effects on signaling pathways related to cancer.

2   Collection and Preparation of Actinomycete Strains

2.1   Collection of Field Samples for Isolation of Actinomycete Strains Principally from the Chiba Area of Japan

Field samples for the isolation of actinomycete strains were collected in various locations in Japan, particularly in the Chiba prefectural area (Fig. 1). Soils, sand from the seaside, and sea water samples were collected in almost 250 locations.

Soil and sand from the seaside samples (1 g) were mixed with sterilized water (10 cm^3), and shaken for 30 min at 60 °C. The supernatant was diluted serially 10 and 100 times with sterilized water and applied to agar plates containing HV or oatmeal media (*7*) (Fig. 2). Colonies of actinomycetes were formed on the media after several days. Shaking at 60 °C was useful for the selective isolation of
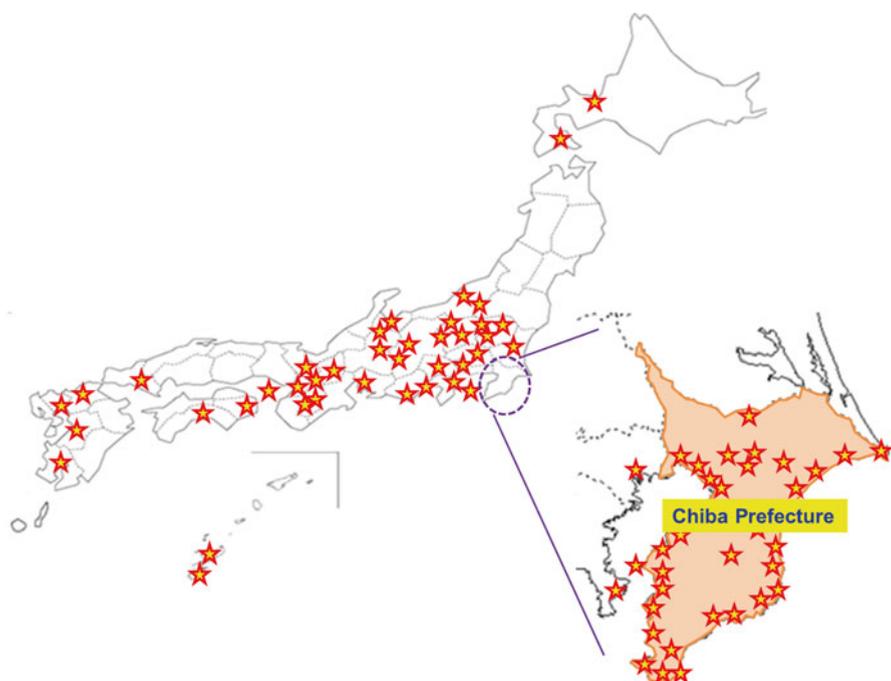
Fig. 1 Collection of field samples from different locations in Japan
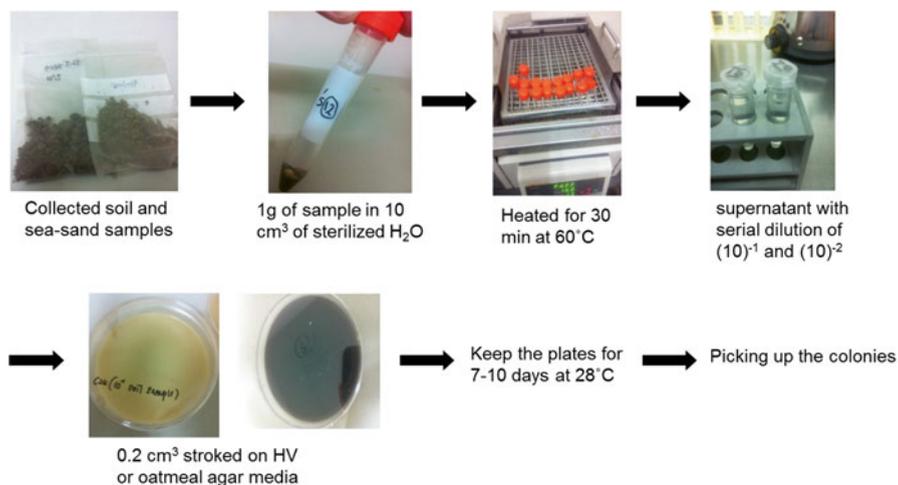
Fig. 2 Isolation of actinomycetes from soil and sand samples

actinomycetes from fungi or other bacteria since actinomycetes are relatively tolerant to heat. Each solution that was diluted ten times gave small and densely crowded aggregate colonies, which were difficult to separate into single colonies. Thus, solutions diluted 1,000 times were used simultaneously.

Sea water samples were submitted to filtration with a nitrocellulose membrane filter (0.22 μm, Millipore), and the filters on which bacteria were attached were placed on the HV agar medium prepared with 50% artificial sea water (2 g of Marine Art SF-1 (Senju Pharmaceutical Co., Ltd.) in 100 cm^3 of sterilized water). The membrane was placed on the HV medium for 5–7 days at 28 °C. The branched mycelia of the actinomycetes penetrated the filter to the underlying agar medium and were able to be separated from non-actinomycete bacteria, which were restricted to the filter surface. The membrane filter was removed and the agar medium was re-incubated for another 7 days at 28 °C to allow for the development of the separated actinomycete colonies (8) (Fig. 3). Using another method, sea water samples were diluted 10 and 100 times with sterilized water and applied directly to the HV agar medium in the same way as performed for the soil and sea-sand samples.

Colonies developed by these two methods on the HV agar medium were inoculated using a platinum loop or a sterilized toothpick on *Waksman* agar plates, and were incubated at 2 °C, with these procedures repeated to give single-strain colonies. Single colonies obtained as described above were inoculated to 1 cm^3 of a 15% glycerol solution in 1.5 cm^3 microtubes to afford glycerol stocks, which were kept at −30 °C as a glycerol stock assembly. Using these methods, 1,245 actinomycete strains have been isolated to date, and these were subjected to small-scale culturing for the preparation of an in-house assembly of actinomycete extracts.

nitrocellulose membrane   bacteria attached on filter   membrane is placed on HV media for 5-7 days at 28°C   take out the membrane filter

Keep the plates for another 7 days at 28°C   picking up the colonies   different strains on Waksman agar plates   Small scale culture on Waksman media
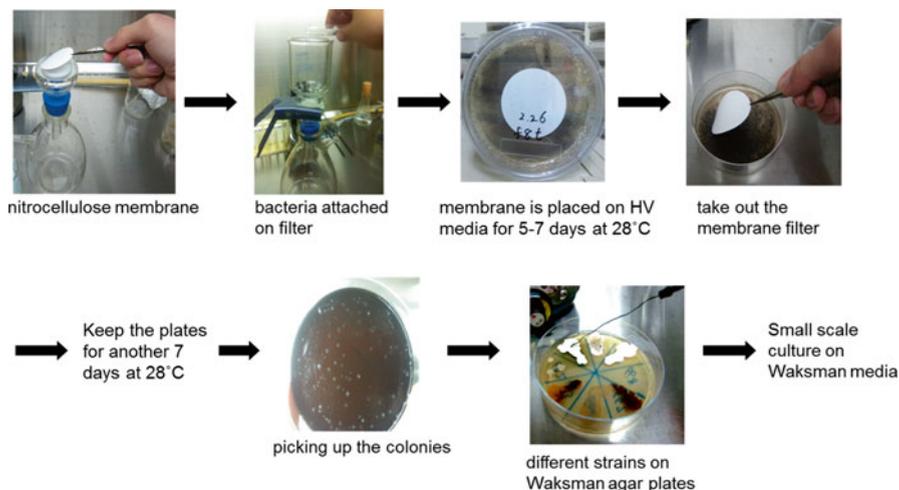
Fig. 3 Isolation of actinomycetes from sea water samples

2.2 Preparation of an Actinomycete Extract Assembly

Single strains developed on *Waksman* agar plates were inoculated on Waksman liquid medium (2 cm^3) in test tubes (Fig. 4), and incubated under shaking (120 rpm) at 28 °C. After five days of incubation, each whole culture was extracted with the same volume of methanol, stirred using a vortex mixer for 1 min, and then subjected to centrifugation (3,000 rpm, 15 min) to separate the supernatant from the mycelium residue. The supernatant containing the methanol extract of the mycelium cake was evaporated under reduced pressure to give a residue, to which 50% ethanol was added to give a solution by adjusting the concentration to 50 mg/cm^3. These 50% ethanol solutions of extracts of actinomycete strains at a concentration of 50 mg/cm^3 were stored in microtubes at −30 °C, and used for various screening tests in our laboratory, as described below.

colonies developed on
Waksman agar plates

Inoculation to *Waksman*
liquid medium

2 cm^3 culture
in test tube

Fig. 4 Small-scale culture of isolated actinomycetes to be used as a screening sample

3 Screening Studies Using the Actinomycete Extract Assembly

3.1 Chemical Screening

The actinomycete extracts that were prepared as described above were then used for various screening studies as follows. First, chemical screening tests using thin-layer chromatography (TLC) were carried out. Different actinomycete crude extracts were applied to silica gel TLC plates using chloroform and methanol as solvents (Fig. 5). Bands of interest were localized under UV light (254 and/or 366 nm) and by using different visualization reagents, such as anisaldehyde, sulfuric acid, phosphomolybdic acid, *Dragendorff*'s reagent, and others. Interesting actinomycete strains selected in this manner were subjected to further scale-up culturing and work up for chromatographic fractionation procedures. Strains producing izumiphenazines, izuminosides, and katorazone, as described in the following sections, were chosen on the basis of TLC-based screening.

414 429 120 323 406 376 386 375 398 367   32   384 385   236 190 240 192 95   25   346 64 373 227   91   179   150 200 146
                    strain numbers   strain numbers

Fig. 5 Silica gel TLC examination of different actinomycete crude extracts using a solvent system of chloroform and methanol (90:10) visualized by anisaldehyde

3.2 Cytotoxicity Testing

As a second process for the selection of interesting actinomycete strains, cytotoxicity testing was carried out against human gastric adenocarcinoma (AGS) cells. The actinomycete extracts prepared above were subjected to such tests at concentrations of 10 and 100 μg/cm^3 using a fluorometric microculture cytotoxicity assay (FMCA) procedure (9). Using this method, fluorescein diacetate was used as a substrate, which was converted into fluorescein by an esterase enzyme present in the living cells, and fluorescence was measured in a 96-well scanning spectrofluorometer at 538 nm, following excitation at 485 nm (Fig. 6).

fluorescein diacetate   fluorescein

Fig. 6 Fluorometric microculture cytotoxicity assay (FMCA) method

Those extracts exhibiting cell viability lower than 60% were judged as active samples. Fractionation of active compounds based on this screening method is described in later sections in this chapter. Many well-known compounds, however, including nocardamine, nonactin, chromomycins, actinomycins, cholic acid, diketopiperazines, and others, were identified on the basis of this method.

3.3 Screening Tests Targeting Signaling Pathways Related to Cancer Disease

3.3.1 TRAIL Signaling (1): Enhancement of Death-Receptor 5 Promoter Activity

Recently, tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL), a member of the TNF superfamily, has become of special interest as a promising target for cancer therapy, since it triggers apoptosis in a variety of cancer cells, but

Fig. 7 Apoptosis-inducing pathway triggered by TRAIL signaling

not in many normal cells (*10*). TRAIL is one of the death ligands and is known to bind to death receptors such as DR5 (death receptor 5) or DR4 (death receptor 4) resulting in the activation of caspase-signaling pathways leading to apoptosis (Fig. 7).

Although TRAIL is a promising target for the discovery of new anticancer agents, recent studies have shown that some cancer cells are intrinsically or acquired resistant to apoptosis induced by TRAIL, which poses a potential restriction to its use. Therefore, for the clinical use of TRAIL in cancer therapy, it is extremely important to overcome TRAIL resistance. TRAIL resistance has been attributed to the loss of TRAIL receptors, upregulation of TRAIL decoy receptors, enhanced expression of cellular FLICE-like inhibitory protein (cFLIP) and cellular inhibitor of apoptosis protein (cIAP), and/or alterations in the expression of the Bcl-2 family of proteins. Recently, it has been reported that TRAIL-resistant cancer cells can be sensitized by combined treatment with TRAIL and cancer chemotherapeutic drugs or some natural products, such as PS-341 (bortezomib), tunicamycin, and curcumin (*11*).

In many cases, overcoming the TRAIL-resistance by combination treatment has included up-regulation of death receptors, especially DR5. Therefore, compounds enhancing DR5 expression are considered to be a new tool to abrogate TRAIL resistance. During studies of our group on the search for bioactive natural products, a screening study has been conducted for substances that enhance the activity of DR5 expression. For this purpose, a luciferase assay system based on a human colon
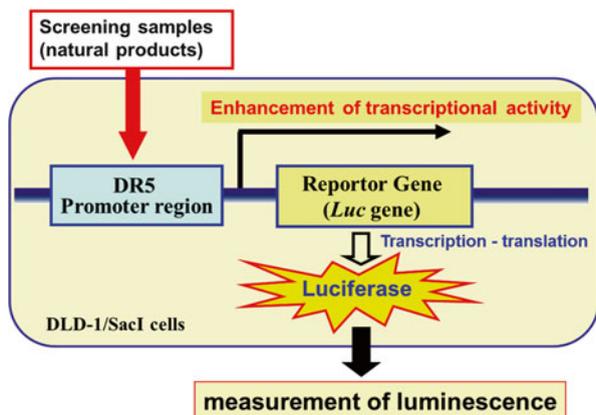
Fig. 8 DR5 promoter activity test

cancer cell line (the DLD-1/*Sac*I cell provided by Professor *Toshiyuki Sakai*, Kyoto Prefectural University of Medicine) has been used, which was stably transfected with a plasmid containing the human DR5 promoter sequence and a luciferase reporter gene (Fig. 8). The up-regulation of the DR5 promoter was assessed by luminescence depending on the luciferase gene expression. DR5 promoter activity has been examined for more than 250 plant extracts in an assembly (*11*, *12*). The species concerned were collected in the northeastern part of Thailand (the Khon Kaen area), and previously the isolation of a series of active compounds, including new isoflavones from *Millettia brandisiana* (Leguminosae) (*13*) and *Ardisia colorata* (Myrsinaceae) (*14*), flavonoids from *Eupatorium odoratum* (Compositae) (*15*), a charcone, cardamomin, from *Catimbium speciosum* (Zingiberaceae) (*16*), and cycloartane triterpenoids from *Combretum quadrangulare* (Combretaceae) (*17*) and *Euphorbia neriifolia* (Euphorbiaceae) (*18*) were reported, including an evaluation of their effects on apoptosis inducement of cancer cells mediated by the TRAIL signaling pathway. The results of a similar screening study by our group using actinomycetes extracts are described in Sect. 9.1.

3.3.2 TRAIL Signaling (2): TRAIL-Resistance Overcoming Activity

Another approach involved with the TRAIL signaling pathway that is being investigated in our group, using a screen for natural products having a reversal effect on TRAIL resistance. Thus, natural products having synergistic effects with TRAIL against TRAIL-resistant tumor cells have been targeted. Through the use of TRAIL-resistant gastric cancer cells (AGS) or TRAIL-resistant DLD1 human colon cancer cells (DLD1/TR), cell viability was compared in the presence and absence of TRAIL by the cytotoxicity test using the FMCA method (Sect. 3.2, Fig. 6). Active samples are those having strong cytotoxicity (= low cell viability) in the presence of TRAIL, compared with the value obtained in the absence of TRAIL (= sample only)
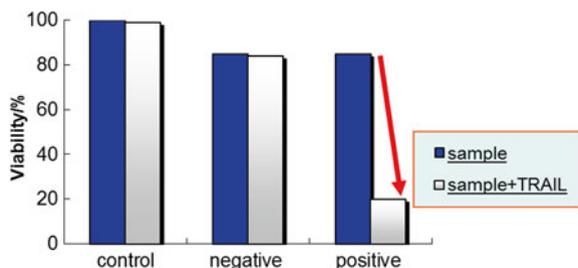
Fig. 9 TRAIL-resistance overcoming activity test

(Fig. 9). In this screening study, more than 200 plant extracts were examined, and several leads were selected with their active constituents pursued, leading to reports of the isolation of several active natural products, including diterpenoids from *Erythrophleum succirubrum* (Leguminosae) (*19*), cardenolide glycosides from *Thevetia peruviana* (Apocynaceae) (*20*), flavonoid glycosides from *Solanum verbascifolium* (Solanaceae) (*21*), rocagloic acid derivatives from *Amoora cucullata* (Meliaceae) (*22*), and sesquiterpenoids from *Kandelia candel* (Rhizophoraceae) (*23*), along with their effects on mRNA or protein expression related to apoptosis inducement mediated by the TRAIL signaling pathway. Results of a similar screening study of the ability to overcome TRAIL resistance using extracts of actinomycetes strains are described in Sects. 4.3, 5.4, 6.3, and 9.2.

3.3.3 Wnt Signaling

Wnt signaling is implicated in the regulation of diverse cellular processes, including cell proliferation, survival, migration, polarity, specification of cell fate, and self-renewal in stem cells. Perturbation of Wnt signaling can result in defects in the development of a cardiovascular, central nervous system, renal, lung, and bone organ systems, or lead to the formation of tumors when activated aberrantly (*24*, *25*). Molecular studies have shown activating mutations of the Wnt signaling pathway to be the cause of approximately 90% of colorectal cancers, in addition to other cancers, such as hepatocellular carcinoma and breast cancer. In the absence of Wnt ligands, β-catenin is phosphorylated by a destruction complex that contains the scaffolding proteins axin and adenomatous polyposis coli (APC), and glycogen synthase kinase 3β (GSK3β). Phosphorylated β-catenin is recognized by the E3 ubiquitin ligase β-TrCP and targeted for proteasomal degradation. Abnormal regulation of Wnt/β-catenin signaling and the subsequent up-regulation of β-catenin expression is a hallmark of the development of certain cancers. Excessive accumulation of cytosolic β-catenin leads to translocation to the nucleus, where this protein functions as a co-factor for transcription factors of T-cell factor/lymphoid enhancing factor (TCF/LEF), and leads to the stimulation of target genes including the c-myc and cyclin D1 genes (Fig. 10). Thus, Wnt/β-catenin signaling is an attractive
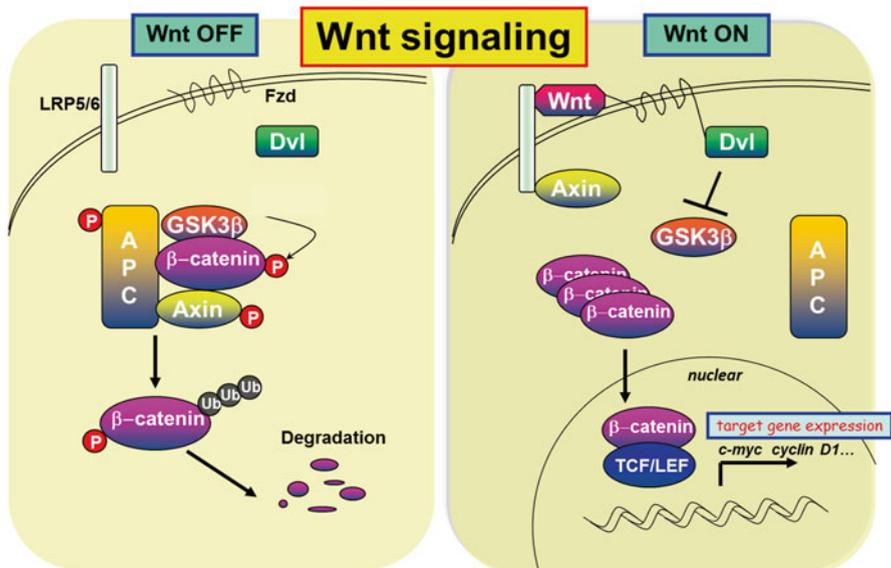
Fig. 10 Wnt/β-catenin signaling pathway

target for cancer therapy, and there is significant interest in the search for Wnt/β-catenin signaling inhibitors and, of these, several small-molecule synthetic and natural products have been recently reported (*26*).

For a screening study on Wnt signaling inhibitors, our group has used a cell-based reporter luciferase assay system to evaluate the inhibition of TCF/β-catenin transcriptional activity. TCF/β-catenin transcriptional activity was measured using the STF/293 cell line (provided by Prof. *J. Nathans*, John Hopkins Medical School, Baltimore, Maryland), which is a 293 human embryonic kidney cell line stably transfected with SuperTopflash, a β-catenin-responsive luciferase reporter plasmid containing eight copies of the TCF-binding site (CCTTTGATC) (provided by Prof. *R. T. Moon*, University of Washington, Seattle, Washington). The 293 cells have a low TCF/β-catenin transcriptional activity due to low endogenous levels of the β-catenin protein. However, Wnt signaling in the cells can be stimulated by inhibiting GSK3β. LiCl, an inhibitor of GSK3β, causes the accumulation of non-phosphorylated β-catenin and, as a result, an increase in TCF/β-catenin transcription. Therefore, STF/293 reporter cells combined with LiCl treatment have been used to evaluate SuperTopflash activity. Extracts inhibiting SuperTopflash activity were further investigated for SuperFopflash activity (luciferase reporter plasmid containing eight copies of a mutant TCF/LEF-binding site (CCTTTGGCC)), in order to exclude non-selective inhibitors of Wnt signaling (Fig. 11).

As a result of using this screening system, more than 300 plant extracts have been examined, and several leads were selected for purification of the active compounds present. Consequently, a number of compounds have been isolated with
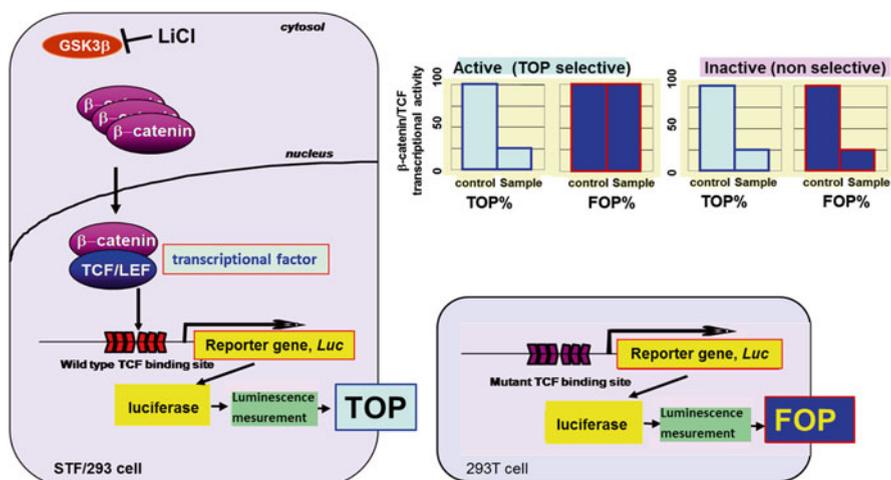
Fig. 11 Cell-based reporter luciferase assay system for Wnt signal inhibition activity

Wnt signal inhibitory activity, including naphthalene quinones from *Eleutherine palmifolia* (Iridaceae) (*27*) and *Impatiens balsamina* (Balsaminaceae) (*28*), and flavonoids from *Bauhinia malabarica* (Caesalpiniaceae) (*29*). In addition, diterpene constituents from *Excoecaria indica* (Euphorbiaceae) were found with Wnt signal augmenting effects (*30*). Recent results on screening studies for Wnt signal activity using extracts of actinomycete strains are described in Sects. 4.3, 6.3, 7.1, and 9.3.

The following five sections (Sects. 4–8) describe recent results on the isolation of new heterocyclic compounds obtained from actinomycete extracts by our group, based mainly on chemical screening methods, as outlined in Sect. 3.1. The compounds to be described were revealed to have effects on the TRAIL or Wnt signal pathways.

4 Izumiphenazines and Izuminosides

4.1 Izumiphenazines

Phenazines are heterocyclic compounds that are produced naturally by various microorganisms, and many of them are known to exhibit a broad range of biological activities, such as antibacterial, antimalarial, antitumor, and antiparasitic effects (*31*). The first phenazine derivative isolated from *Streptomyces* sp. was the antibiotic griseolutein (*32*), and since then, an increasing number of phenazine derivatives with different activities have been isolated from different Streptomycete species such as phenacein (*33*), phenazinomycin (*34*), phenazostatin A (*35*), and others (Fig. 12).
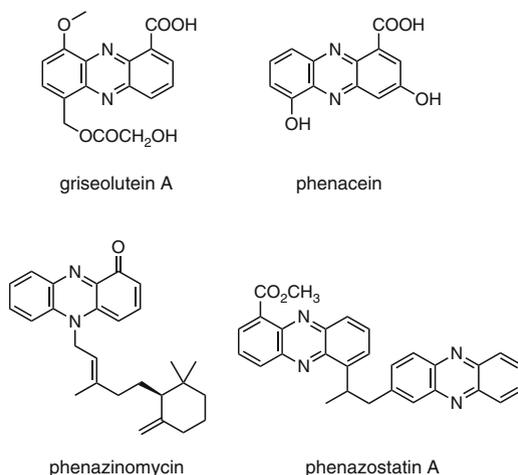
Fig. 12 Examples of phenazine compounds obtained from Streptomycetes

In the course of screening for bioactive compounds from natural sources, an extract of *Streptomyces* sp. IFM 11204 was investigated because it exhibited relatively polar yellow bands by TLC that turned orange upon spraying with anisaldehyde reagent, and dark red with concentrated sulphuric acid, and reddish brown with *Dragendorff*'s reagent. This and the lack of a color reaction with sodium hydroxide pointed to the presence of compounds bearing a phenazine skeleton. The strain *Streptomyces* sp. IFM 11204 was isolated from a soil sample collected from Izumi Forest, Chiba, Japan. Fermentation was carried out at 28 °C for 5 days while shaking at 200 rpm. The seed culture (10 cm^3) was used to inoculate ten 3-dm^3 flasks each containing 500 cm^3 of the same medium, which were incubated using similar conditions. After centrifugation of the culture medium (5 dm^3), extraction with acetone, and evaporation, the crude extract was fractionated by several chromatographic processes to result in the isolation of the three new compounds, izumiphenazines A-C (**1–3**), along with the known phenazine-1,6-dicarboxylic acid (**5**) (Fig. 13) (*36*). From another culture (14 dm^3) of *Streptomyces* sp. IFM 11204, 20 3-dm^3 flasks each containing 700 cm^3 of the same medium, izumiphenazine D (**4**) and three known phenazine compounds (**6–8**) (Fig. 13) (*37*) were obtained by fractionation using silica gel, octadecylsilane, and Sephadex LH-20 chromatographic procedures.

Izumiphenazine A (**1**) gave a molecular formula of C$_{25}$H$_{16}$N$_4$O$_6$, as suggested from its HRESIMS data, and its ^{1}H NMR spectrum in DMSO-d_6 showed signals due to three hydroxy group protons, seven sp^2 methines, three sp^3 methines, and one sp^3 methylene group. Interpretation of the 2D-NMR spectra of **1** revealed the presence of two phenazine units, namely, a 3,6-dihydroxy-1-phenazine carboxylic acid (unit A) and a 1,2,3,4-tetrahydrophenazine-1,9-diol (unit B). These units (A and B) were connected through a C-C bond between C-4 and C-3′ and an ether oxygen between C-3 and C-2′ (Fig. 14).
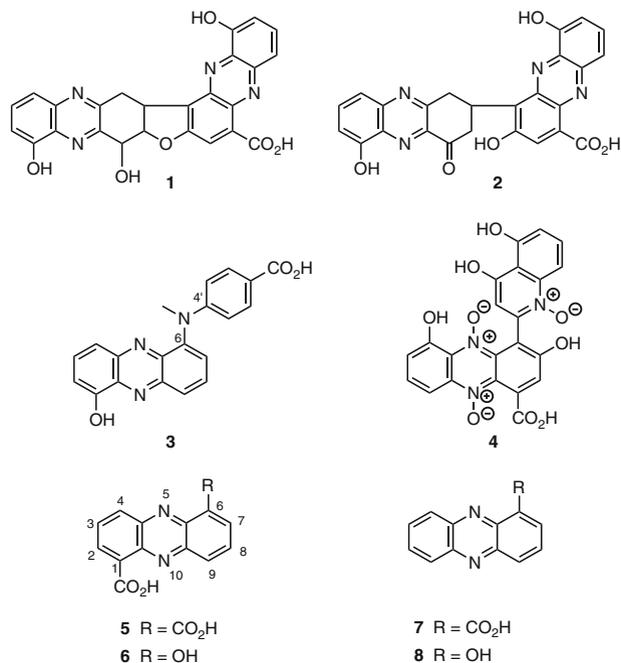
Fig. 13 Izumiphenazines A-D (**1–4**) and known phenazine compounds (**5–8**)
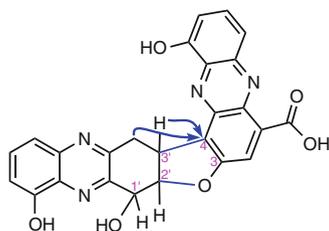
Fig. 14 Structure elucidation of izumiphenazine A (**1**)

Izumiphenazine B (**2**) gave an identical molecular formula to izumiphenazine A (**1**), and was found to contain the same phenazine component, unit A. Instead of the unit B of izumiphenazine A (**1**), izumiphenazine B (**2**) was shown to possess a 9-hydroxy-1,2,3,4-tetrahydrophenazin-1-one unit, of which the C-3′ position was connected at C-4 of unit A by a C—C bond (Fig. 15).

Izumiphenazine C (**3**), with a molecular formula of $C_{20}H_{16}N_3O_3$, showed signals for ten sp^2 methines as well as an *N*-methyl group in its ^1H NMR spectrum. HMBC correlations were observed from the *N*-methyl protons to C-6 and C-4′, indicating that compound **3** has a 1-hydroxyphenazine unit and a benzoic acid moiety, which were connected through an *N*-methyl group (Fig. 16).
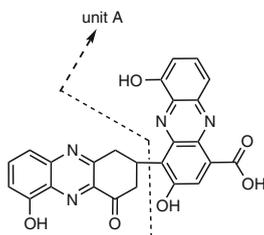
Fig. 15 Structure elucidation of izumiphenazine B (**2**)
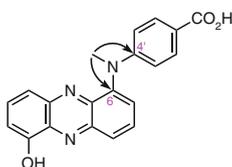
Fig. 16 Structure elucidation of izumiphenazine C (**3**)
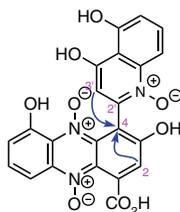
Fig. 17 Structure elucidation of izumiphenazine D (**4**)

Izumiphenazine D (**4**) was assigned a molecular formula of $C_{22}H_{13}N_3O_9$ based on its HRESIMS data, and spectroscopic data analysis revealed that this compound consists of phenazine *N*-dioxide and quinolone *N*-oxide units. These were connected by a C-C bond at the C-4 and C-2′ positions, based on the HMBC correlations observed from both H-2 and H-3′ to C-4 (Fig. 17).

Several phenazine *N*-dioxides such as iodinin and lomofungin (*31*) were previously known from different actinomycetes, while many quinolone *N*-oxide alkaloids like KF8940 (*38*) and aurachin A (*39*) have been isolated from a number of microbial sources (Fig. 18). However, to the best of our knowledge, izumiphenazine D (**4**) is the first example of a microbial natural product containing both a phenazine *N*-dioxide and a quinolone *N*-oxide unit.
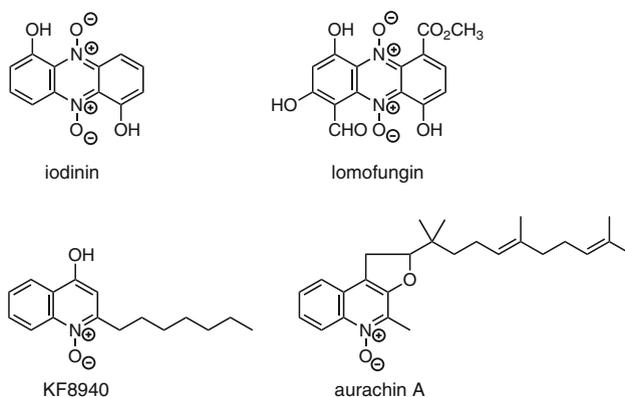
Fig. 18 Previously known phenazine *N*-dioxides and quinolone *N*-oxides

4.2 *Izuminosides*

Phenazines comprise a large group of nitrogen-containing heterocyclic compounds. More than 100 different phenazine structural derivatives have been identified from natural sources, and over 6,000 compounds that contain phenazine as a central moiety have been synthesized. However, carbohydrate-containing phenazine natural products are rare in Nature (*31*). From literature, phenazoviridine was the first glycoconjugated phenazine obtained from the actinomycetes (*40*). Other examples of carbohydrate-containing phenazines are aestivopheonins A-C (*41*), and the 2-*O*– and 3-*O*-L-quinovosyl esters of saphenic acid (*42*) (Fig. 19).

From a soil sample, again collected from Izumi Forest, Chiba, Japan, *Streptomyces* sp. IFM 11260 was isolated. TLC examination of a crude extract of this strain showed several relatively polar yellow bands, which were not visible under UV light (254 nm). These bands gave a dark red color with concentrated sulphuric acid and turned to reddish brown after spraying with *Dragendorff*'s reagent. Thus, this *Streptomyces* strain was cultured in *Waksman* medium at 28 °C for 5 days on a rotary shaker. After centrifugation, extraction, and solvent evaporation, work-up of the crude extract resulted in the isolation of three new glycoconjugated phenazine alkaloids, designated as izuminosides A-C (**9–11**) (Fig. 20) (*43*).

The molecular formula of izuminoside A (**9**) was determined as $C_{19}H_{18}N_2O_7$ by HRESIMS. Analysis of its 2D-NMR spectra suggested that compound **9** consists of a 6-hydroxyphenazine-1-carboxylic acid aglycone and a 6-deoxyhexose unit. The aglycone part was confirmed by comparison of the spectroscopic data with the literature values, and was found to be identical with the known compound **6** isolated from *Streptomyces* sp. IFM 11204 (Fig. 13). The 6-deoxyhexopyranose units in compound **9** was identified as rhamnose from the N*O*E data and coupling constants (Fig. 21). The N*O*E observed between H-3′ and H-5′ as well as the coupling constant
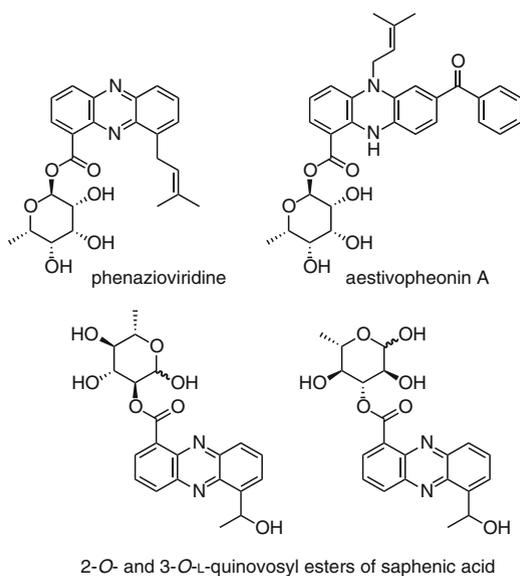
Fig. 19 Examples of carbohydrate-containing phenazine natural products
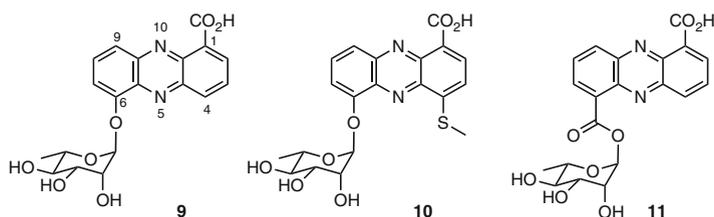
Fig. 20 Izuminosides A-C (**9–11**)
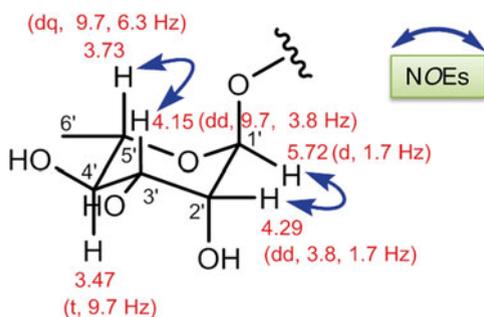
Fig. 21 Selected N*O*E correlations and *J* values of the rhamnose residue of **9**

between H-4′ and H-5′ ($J_{4',5'} = 9.7$ Hz) indicated that these protons are *axially* oriented. The configuration of the anomeric proton of compound **9** was assigned as α based on the small coupling constant ($J_{1',2'} = 1.7$ Hz). The anomeric configuration was also deduced as α from the one-bond coupling constant between C-1′ and H-1′ (Rha, $J_{C1',H1'}$ values: 171.9 Hz observed for **9**; literature values, α-anomer, 172 Hz; β-anomer, 160 Hz) (*44*).

The sugar moiety in compound **9** was confirmed as rhamnose by HPLC analysis (Capcell Pak NH_2 UG80, 85% CH_3CN) of the crude sugar obtained by acidic hydrolysis. The absolute configuration of the sugar residue was determined to be L- by comparison with the authentic samples using a combination of RI and optical rotation detectors. The sugar part of the hydrolyzate of **9** was dextrorotatory ($[\alpha]_D^{20}$ +8.4°cm^{2}g^{-1} (*c* 0.13, water)), while an optical rotation detector (JASCO OR-2090) showed a negative peak because the optical rotation detector did not use a single wavelength. Thus, the structure of izuminoside A (**9**) was determined as 6-(1′–*O*-α-L-rhamanopyranosyl)-phenazine-1-carboxylic acid.

The aglycone of izuminoside C (**11**) was found to be a known compound, phenazine-1,6-dicarboxylic acid (**5**), which was also isolated from *Streptomyces* sp. IFM 11204 (Fig. 13), whereas the aglycone of izuminoside B (**10**) was revealed to be an unknown compound, 4-methylsulfanyl-6-hydroxyphenazine-1-carboxylic acid.

4.3   Effects on Wnt and TRAIL Signaling

The isolated compounds were evaluated for their activities against Wnt signal inhibition activity and TRAIL resistance overcoming activity. The assay results (Fig. 22) showed that izumiphenazines A (**1**), B (**2**), and C (**3**) showed moderate inhibition of Wnt signal transcriptional activity with IC_{50} values of 81.4, 24.7, and 84.1 μM.
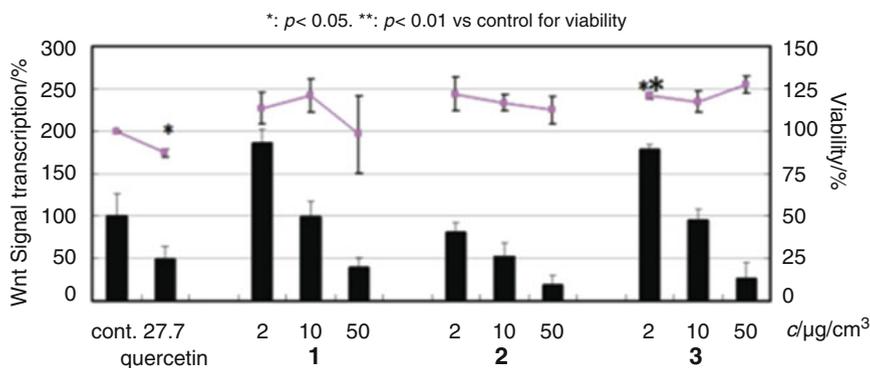
Fig. 22 Wnt signal inhibitory activity of izumiphenazines A-C (**1–3**). The bar chart shows Wnt signal transcriptional activity and the line chart represents the viability of the STF/293 cells. Quercetin was used as a positive control

Compounds **1**–**3** were evaluated for their ability to overcome TRAIL resistance in human gastric adenocarcinoma (AGS) cells. This cell line has been used widely as a model system for evaluating cancer cell apoptosis since it was reported to be refractory to apoptosis induction by TRAIL. To assess the TRAIL resistance overcoming activity, AGS cells were treated with compounds **1**–**3** in the presence and absence of TRAIL, and the cell viability was examined by the fluorometric microculture cytotoxicity assay (FMCA) method. Luteolin (lut) was used at 18.0 μ*M* as a positive control (*45*). The assay results (Fig. 23) showed that compounds **2** (30 μ*M*) and **3** (20 μ*M*) exhibited 30 and 26% decreases in cell viability, in the presence of TRAIL (100 ng/cm^3) compared with in the absence of TRAIL. On the other hand, treating the cells with compound **1** (40 μ*M*) at 100 ng/cm^3 TRAIL reduced the cell viability to 78%, which was 19% more than by TRAIL alone. These results suggest that izumiphenazines A-C (**1**–**3**) showed a synergistic effect in combination with TRAIL against AGS cells.
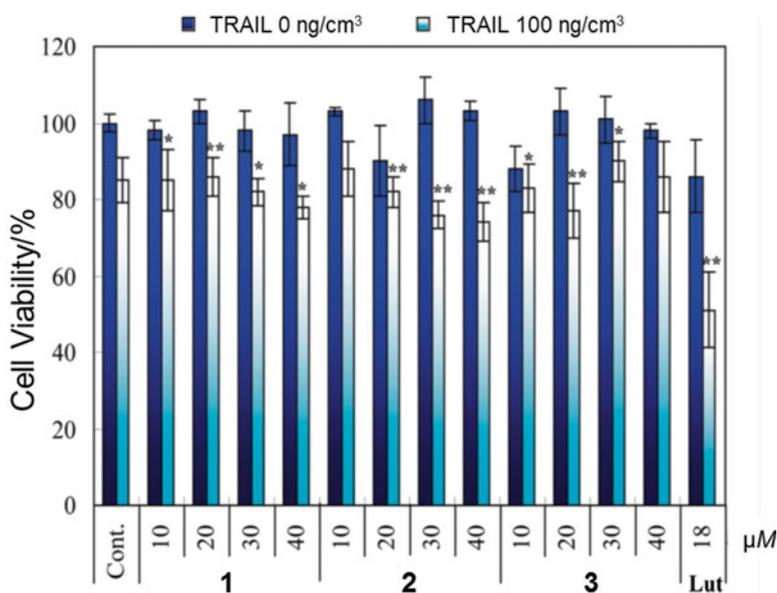
Fig. 23 Effect of compounds **1**–**3** on the cell viability of AGS cells in the presence and absence of TRAIL. Luteolin was used as a positive control

The resistance of cancer cells toward TRAIL may occur at different points in the TRAIL-induced apoptotic pathway. Understanding the mechanisms of such resistance and developing strategies to overcome it are important for the successful use of TRAIL in cancer therapy. Combined treatment with TRAIL and chemotherapeutic agents, including natural products, can overcome such resistance and sensitize TRAIL-resistant cells to enhance the therapeutic potential of TRAIL against cancer cells. Therefore, a natural product that elicits synergistic activity with TRAIL would be a new tool for investigating cancer cells. Izumiphenazines A-C (**1**–**3**) were, to the

best of our knowledge, the first phenazine dimers having synergistic activity in sensitizing TRAIL-resistant AGS cells, thereby suggesting their possible use in combination with TRAIL against human gastric adenocarcinoma.

Izumiphenazine D (**4**) was also found to be moderately active in this assay. At a 40 μ*M* concentration of **4**, the cell viability of AGS cells was 99%, whereas the same concentration of this substance in the presence of 100 ng/cm^3 of TRAIL reduced cell viability to 80% of control levels, which was 19% lower than compound **4** alone. Wnt signal inhibitory activity of izumiphenazine D (**4**) was also examined using a luciferase reporter gene assay in SuperTOP-Flash transfected cells to show that compound **4** did not exhibit inhibition of Wnt signal transcription activity even at 50 μ*M*.

The bioactivities of izuminosides A-C (**9**–**11**) were evaluated in terms of their ability to overcome TRAIL resistance in AGS cells. To assess the effects of compounds **9**–**11** on cell viability, in the presence and absence of TRAIL, AGS cells were treated with these compounds and then subjected to evaluation using the FMCA method. As a result, compounds **10** (10 μ*M*) and **11** (60 μ*M*) exhibited 22 and 19% decreases in cell viability in the presence of TRAIL (100 ng/cm^3), when compared to the absence of TRAIL. Compound **9**, however, did not show TRAIL-resistance overcoming activity. The results obtained suggested that izuminosides B (**10**) and C (**11**) gave a synergistic effect in combination with TRAIL using AGS cells.

5   Pyranonaphthoquinones

5.1   *Pyranonaphthoquinones*

Pyranonaphthoquinones are a class of compounds that contain a core naphtha[2,3-*c*] pyran-5,10-dione ring system (Fig. 24), and have been isolated from various organisms such as plants, fungi, and actinomycetes (*46*).

Eleutherin is a representative compound of this group, which was isolated from Iridaceous plants (*47*), and shown to possess the basic pyranonaphthoquinone skeleton with little functionalization. An additional γ-lactone is attached in kalafungin, and nanaomycin A has an open-chain form of the γ-lactone. Griseusin A possesses an additional dihydropyran ring with spiro structure, and cardinalin C3 has a dimeric pyranonaphthoquinone structure with atropisomerism (Fig. 25). Most of these pyranonaphthoquinones are of microbial origin and are known to possess various biological effects such as antibacterial, antiviral, and cytotoxic activities (*46*). In this connection, eleutherin was isolated from *Eleutherine palmifolia* (Iridaceae) and it was found that this compound as well as its related naphthalene quinones and their glycoside exhibit Wnt signal inhibitory activity (*27*).

In a collection of actinomycetes obtained by our group, which were isolated from soil and sea water samples from different areas of Japan, a crude extract of the
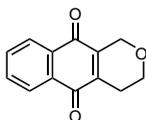
Fig. 24 The core structure of pyranonaphthoquinone

eleutherin   kalafungin   nanaomycin A

griseusin A   cardinalin C3
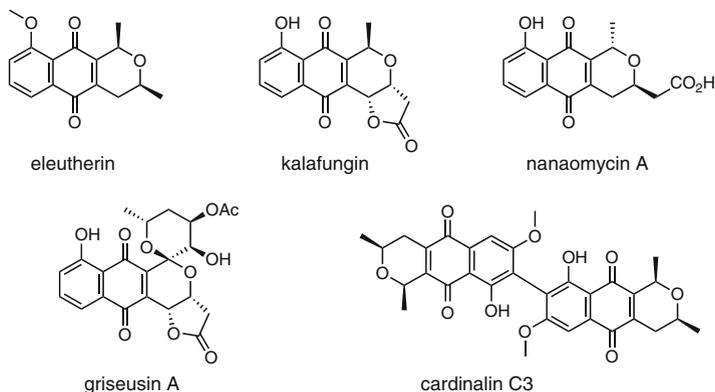
Fig. 25 Examples of pyranonaphthoquinone natural products

terrestrial *Streptomyces* sp. IFM 11307 drew attention due to its striking yellow bands on TLC, which gave a blue color reaction with $2N$ NaOH and a strong orange fluorescence under UV light (366 nm). This strain, IFM 11307, was isolated in 2007 from a soil sample collected from Yoro Valley, Ichihara-shi, Chiba prefecture, Japan (Fig. 26).

Fig. 26 *Streptomyces* sp. IFM 11307

Well-grown agar cultures of *Streptomyces* sp. IFM 11307 were used to inoculate 4×500 cm³ Sakaguchi flasks, each containing 100 cm³ of Waksman medium. The flasks were incubated at 28 °C while shaking at 200 rpm for 5 days. The seed culture (10 cm³) was used to inoculate 16 3-L flasks, each containing 750 cm³ of the same medium, which were incubated using similar conditions. After centrifugation and extraction of the culture broth (12 L), work-up of the crude extract, followed by fractionation using silica gel, ODS, and Sephadex LH-20 chromatography, this resulted in the isolation of four new spiropyranonaphthoquinones (**12–15**) (Fig. 27) as well as a new phenazine compound (**16**) (*48*), in addition to a known phenazine dicarboxylic acid (**5**), which was previously isolated (*37*) from *Streptomyces* sp. IFM 11204 (Fig. 13).
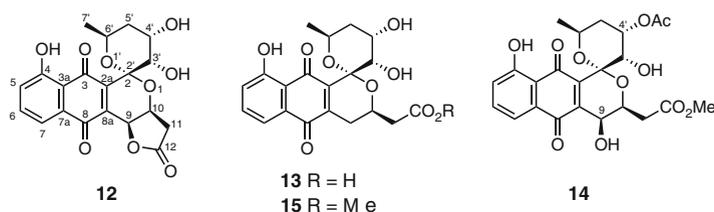
Fig. 27 New spiropyranonaphthoquinone compounds isolated from *Streptomyces* sp. IFM 11307

Compounds **12** and **13** were both isolated as yellow solids. The molecular formula of **12** was established as $C_{20}H_{18}O_9$ by HRESIMS (*m/z* 425.0811 [M+Na]⁺, Δ −3.7 mmu), and compound **13** gave a molecular formula of $C_{20}H_{20}O_9$, as revealed by HRESIMS (*m/z* 427.0992 [M+Na]⁺, Δ −1.3 mmu). The spectroscopic data including NMR and MS of compounds **12** and **13** were identical to those of (−)-4′-deacetylgriseusin A and B (*49*), respectively, which indicated the compounds to have the same relative stereochemistry. An interesting observation was that compounds **12** and **13** were found to possess opposite signs of optical rotation values and CD spectra to those of (−)-4′-deacetylgriseusins A and B. Therefore, compounds **12** and **13** were suggested to be the (+)-enantiomers of the previously reported (−)-4′-deacetylgriseusins A and B, respectively. It is well known that kalafungin (*50*) and nanaomycin D (*51*), which are pyranonaphthoquinones similar to **12** or **13**, have an identical relative stereochemistry with the opposite absolute configurations (Fig. 28).

Compound **14** was isolated as yellow solid and gave an orange fluorescence under UV light at 366 nm and a blue color reaction with $2N$ NaOH. The molecular formula was determined as $C_{23}H_{24}O_{11}$ from HRESIMS (*m/z* 499.1227 [M+Na]⁺, Δ +1.1 mmu). Based on the analysis of the 2D-NMR spectra (¹H–¹H COSY, HMQC and HMBC) of **14**, the structure of this compound was elucidated as the 9-hydroxy-4′-*O*-acetyl derivative of compound **13** (Fig. 27). Its relative configuration was
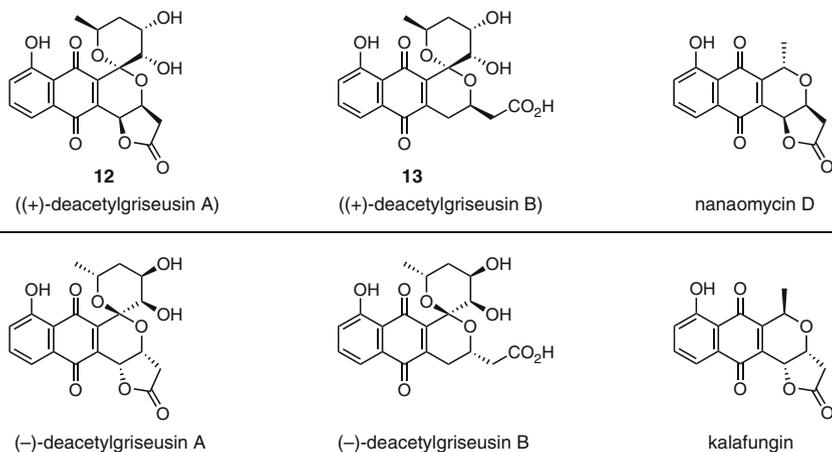
Fig. 28 Enantiomers of pyranonaphthoquinones

12
((+)-deacetylgriseusin A)

13
((+)-deacetylgriseusin B)

nanaomycin D

(−)-deacetylgriseusin A

(−)-deacetylgriseusin B

kalafungin
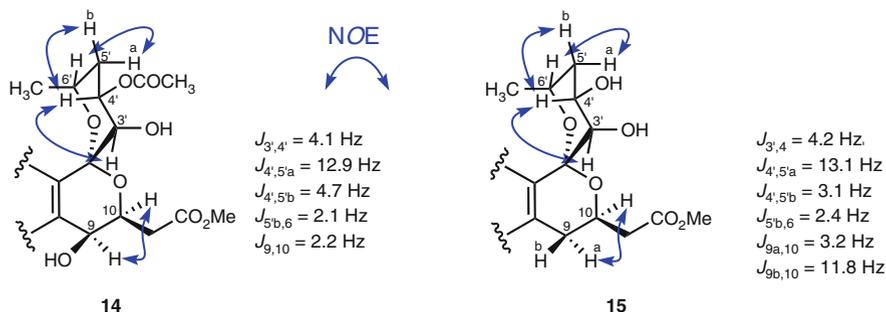
Fig. 29 Selected N*O*E and coupling constants of compounds **14** and **15**

$J_{3',4'} = 4.1$ Hz
$J_{4',5'a} = 12.9$ Hz
$J_{4',5'b} = 4.7$ Hz
$J_{5'b,6} = 2.1$ Hz
$J_{9,10} = 2.2$ Hz

$J_{3',4} = 4.2$ Hz
$J_{4',5'a} = 13.1$ Hz
$J_{4',5'b} = 3.1$ Hz
$J_{5'b,6} = 2.4$ Hz
$J_{9a,10} = 3.2$ Hz
$J_{9b,10} = 11.8$ Hz

14

15

determined from N*O*E and coupling constant observations (Fig. 29) (*48*, corrigendum). The N*O*E observed between the *vicinal* protons H-3′ and H-4′ supported the configurations at C-3′ and C-4′, as depicted in Fig. 29. This relative configuration was supported by the coupling constant of *J*=4.1 Hz between H-3′ and H-4′. Additionally, N*O*E correlations between H-4′ and H-5′b, and H-6′ and H-5′a suggested the configuration of the upper pyran ring in **14** shown in Fig. 29. The *cis* relationship of H-9 and H-10 was confirmed by the N*O*E between H-9 and H-10 ($J_{9,10}$=2.2 Hz). It was found that the molecular formula of compound **15** is $C_{21}H_{22}O_9$ by HRESIMS (*m/z* 441.1151 [M+Na]$^+$, Δ −1.1 mmu). The connectivity

of all protons and carbons was established from the ^1H–^1H COSY, HMQC and HMBC data. The HMBC correlation of the -OCH$_3$ group (δ_H 3.74) to the carbonyl signal at C-12 (δ_C 171.1) confirmed the presence of the methyl ester group in **15**. After comparing these data with literature values (*49*), the constitution of **15** was elucidated as the methyl ester of **13**. The relative configuration of compound **15** was established by analysis of its N*O*E spectra (Fig. 29). Thus, the relative stereochemistry at C-3′, C-4′, and C-6′ was assigned on the basis of the N*O*E correlations between H-3′ and H-4′ ($J_{3',4'} = 4.1$ Hz), H-4′ and H-5′b, and H-6′ and H-5′a.

5.2 Yorophenazine

Compound **16** was obtained as an optically active ($[\alpha]^{22}_D + 47°cm^2g^{-1}$) yellow solid. It gave a positive color reaction with *Dragendorff*'s reagent and fluorescence under UV light at 254 nm. The molecular formula was determined as $C_{20}H_{17}N_3O_7S$ by negative HRESIMS (*m/z* 442.0695 [M−H]$^-$, Δ −1.4 mmu). Compound **16** was identified readily as a phenazine-type compound by considering the UV data (maxima were observed at 449, 370, and 251 nm) and the characteristic low-field chemical shifts of the aromatic protons. Interpretation of the HMBC correlations of **16** suggested the presence of a *para*-substituted phenazine-1,6-dicarboxylic acid unit. Several biosynthetic studies have demonstrated that phenazine-1,6-dicarboxylic acid is a universal precursor for many phenazine secondary metabolites (*52*). The ^1H and ^{13}C NMR chemical shifts for **16** suggested the presence of an amino acid, and these were consistent with a *N*-acetylcysteine methyl ester substituent. Since the aromatic proton H-2 (δ_H 8.32 ppm) and H$_2$-3′ (δ_H 3.84 and 3.49 ppm) showed HMBC correlations to C-4 (δ_C 144.8 ppm), the *N*-acetylcysteine methyl ester was located at the C-4 position. Thus, the structure of **16** was determined as shown in Fig. 30, which was named yorophenazine. From literature, SB 212305 is the only phenazine antibiotic linked to *N*-acetylcysteine found thus far (*53*).

16
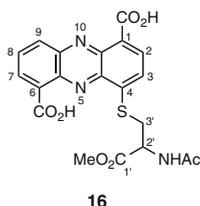
Fig. 30 Yorophenazine (**16**)

5.3 Yoropyrazone

Since the strain *Streptomyces* sp. IFM 11307, isolated from a soil sample collected from Yoro Valley, Chiba, Japan, was found to produce several new bioactive compounds such as pyranonaphthoquinones (**12**–**15**) and yorophenazine (**16**), this strain was investigated further by repeating its culturing in larger scale. A total of 24 dm^3 of the fermentation broth was centrifuged to give the mycelium and the water phase. The resulting mycelial cake was extracted three times with acetone. After removal of acetone, the aqueous solution was extracted three times with EtOAc. The EtOAc-soluble portion was concentrated under reduced pressure. The culture broth was extracted three times with EtOAc. As the TLC of both extracts from the culture filtrate and mycelium showed the same composition, they were combined and concentrated under reduced pressure. The crude extract (7.2 g) was subjected to chromatography over silica gel and Sephadex LH-20 columns, followed by purification with preparative ODS-HPLC, to afford a new compound, yoropyrazone (**17**) (*54*).

Yoropyrazone (**17**) (Fig. 31a) was obtained as a yellow solid, and showed a green fluorescence on TLC under UV light (365 nm) and turned reddish brown after spraying with *Dragendorff*'s reagent. It exhibited a [M−H]$^-$ ion at *m/z* 571.2021 in the HRESIMS, which was consistent with the molecular formula, $C_{27}H_{31}N_4O_{10}$ (Δ −1.9 mmu). Its UV spectrum, with absorption maxima at λ_{max} 408 and 275 nm, was indicative of the presence of a quinone chromophore. The ^{13}C NMR and DEPT spectra of **17**, with the aid of HMQC data, showed the occurrence of 27 carbon signals, and detailed analysis of the 2D-NMR spectra (^{1}H–^{1}H COSY,
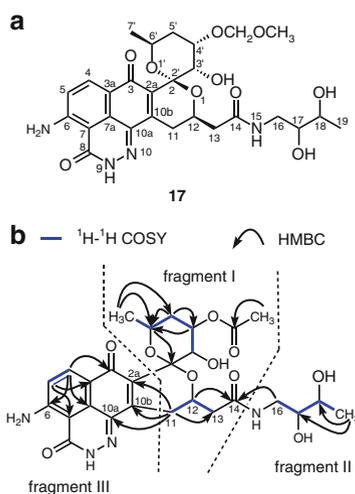
Fig. 31 (**a**) Structure of yoropyrazone (**17**). (**b**) Key COSY and HMBC correlations of yoropyrazone (**17**)

HMQC and HMBC) of **17** led to the assignment of three fragments, I-III (Fig. 31b). The partial structure of fragment I was also contained in pyranonaphthoquinones **12–15** isolated previously from this strain, and fragment II was identified as aminobutane-2,3-diol from the ^{1}H–^{1}H COSY experiments depicted in Fig. 31b. By considering the chemical shift of C-16 (δ_C 43.4 ppm) and the HMBC correlations of H$_2$-16 (δ_H 3.31 and 3.08 ppm) and H-12 (δ_H 4.57 ppm) to the carbonyl group C-14 (δ_C 177.3 ppm), fragments II and I were linked by an amide bond. In fragment III, the HMBC couplings of H$_2$-11 allowed C-2a, C-10a, and C-10b to be assigned. By considering the chemical shift of C-10a (δ_C 133.3 ppm) with the remaining three nitrogen atoms unassigned, fragment III was elucidated as naphthopyrida-zone on the basis of comparison of the chemical shifts of model compounds in the literature (Fig. 32). For the corresponding positions of C-10a of **17**, C-6 of pyrida-zine-3(2*H*)-one resonated at δ_C 139.0 ppm (*55*) and C-8 of the phenylacetic acid hydrazide derivative NG-061 resonated at δ_C 132.8 ppm (*56*). The ^{1}H NMR spec-trum of **1** in DMSO-d_6 solution showed a signal due to the amino group at δ_H 13.5 ppm (2H, broad singlet), which was consistent with the chemical shift value (δ_H 13.2, 2H) of the corresponding amino group of a synthetic compound, 8-amino-2-methyl-2,7-dihydro-3*H*-dibenzo[*de,h*]cinnoline-3,7-dione (*57*). From these observations, the structure of yoropyrazone was concluded as **17**, as shown in Fig. 30a. Some naphthoquinones fused with a pyridazone ring were prepared synthetically and found to exhibit cytotoxic activity (*58*). To the best of our knowl-edge, yoropyrazone (**17**) is the first example of microbial naphthopyridazone natural product.
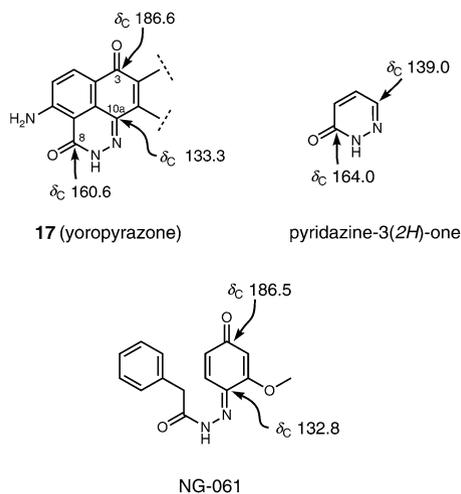
Fig. 32 ^{13}C NMR chemical shift comparison of yoropyrazone (**17**), pyridazine-3(2*H*)-one, and NG-061

5.4   Effects on TRAIL Signaling

The ability to overcome TRAIL resistance of the isolated new spiropyranonaphtho-
quinones **12–15** and yorophenazine (**16**) was examined as follows. To assess the
effects of these compounds on cell viability in the presence and absence of TRAIL,
AGS cells were treated with the indicated agents and were subjected to evaluation
using the FMCA method. Luteolin (lut) was used as a positive control (*45*), produc-
ing about 44% more inhibition along with TRAIL than the agent alone at 17.5 μ*M*.
The assay results (Fig. 33) showed that compound **12** at 0.1 μ*M* proved to be the
most active in this series, with a 33% decrease in cell viability in the presence of
TRAIL (100 ng/cm^3), when compared with the absence of TRAIL. Combined treat-
ment of TRAIL and **13** at 0.5 μ*M* resulted in a 19% greater inhibition than the agent
alone. Compound **14** at 0.5 and 1.5 μ*M* exhibited 20% and 23% decreases, respec-
tively, in cell viability in the presence of TRAIL (100 ng/cm^3), while compound
15 at 0.1 and 0.5 μ*M* caused 28 and 27% decreases in cell viability in the presence
of TRAIL (100 ng/cm^3). Yorophenazine (**16**), however, did not produce any signifi-
cant reduction in cell viability with TRAIL. These results suggested that pyranon-
aphthoquinones **12–15** had a synergistic effect in combination with TRAIL in the
human gastric adenocarcinoma (AGS) cell line.

   The effect of yoropyrazone (**17**) on the viability of human gastric adenocarci-
noma (AGS) cells in the presence and absence of TRAIL was also examined, and
the results are shown in Fig. 34. Compound **17** at 10 μ*M* reduced cell viability to
68% of the control level, whereas combined treatment along with 100 ng/cm^3 of
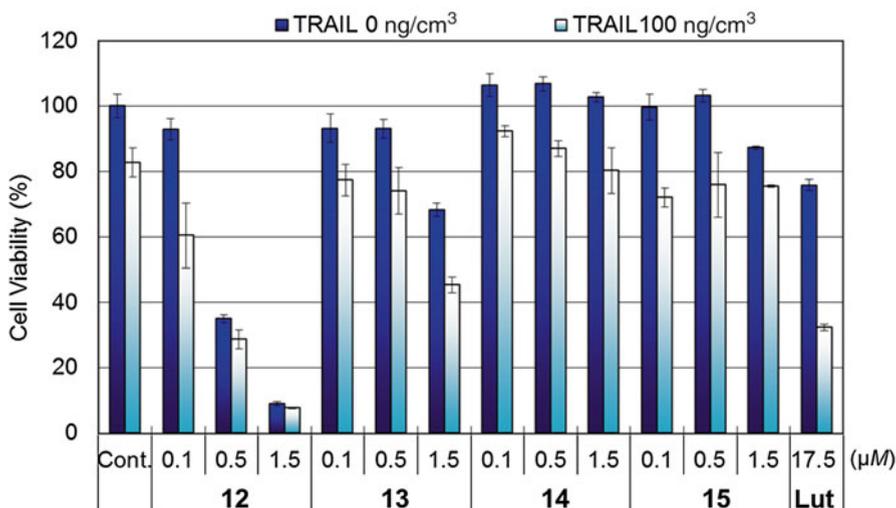
Fig. 33 Effect of pyranonaphthoquinones (**12–15**) on the cell viability of AGS cells in the pres-
ence and absence of TRAIL. Luteolin was used as a positive control
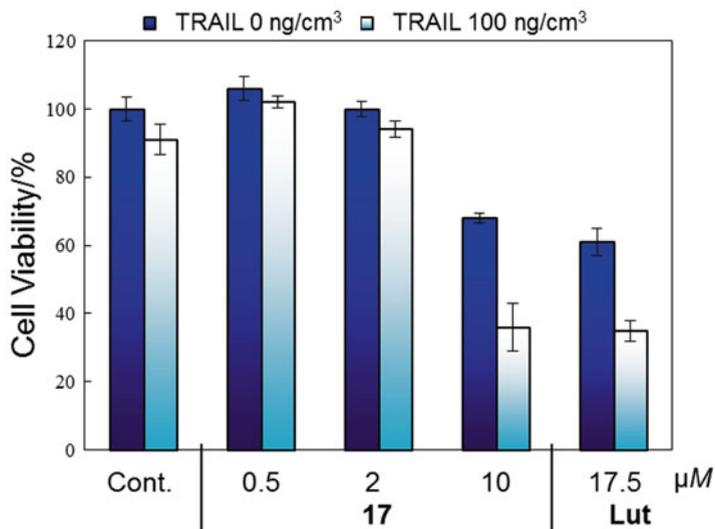
Fig. 34 Effect of yoropyrazone (**17**) on the cell viability of AGS cells in the presence and absence of TRAIL. Luteolin was used as a positive control

TRAIL further reduced the cell viability to 36% of control levels, which was 32% more than the agent alone. Compound **17** also showed cytotoxicity against AGS cells with an IC_{50} value of 12.5 μM. These results suggested that compound **17** in the presence of TRAIL showed a moderate TRAIL-resistance overcoming activity against AGS cells.

6   Azaquinones

6.1   Katorazone

Azaquinone-containing natural products are rarely found in Nature (Fig. 35), having been isolated from fungi (*59*, *60*) and lichens (*61*), and some of them were reported to exhibit antibacterial properties (*62*). Characteristic structural features of naturally occurring 2-azaquinones are the presence of a methyl group at C-3 and different substituents such as OH or OCH_3 groups in ring C. Phenylhydrazones have different uses in organic chemistry including the protection of carbonyl groups (*63*), and the derivatization of natural products (*64*), and increase the sensitivity of carbohydrates for mass spectrometry and ultraviolet detection during high-performance liquid chromatography (*65*).

Fig. 35 Examples of azaanthraquinone natural products

In continuing studies on chemical screening of actinomycete extracts, it was found that the crude extract of *Streptomyces* sp. IFM 11299 showed many yellow bands that fluoresced with an orange color under UV light (366 nm). These bands exhibited a violet color reaction after being sprayed with anisaldehyde or sulphuric acid reagent and one band turned reddish brown after spraying with *Dragendorff*'s reagent. This strain of S*treptomyces* sp. IFM 11299 was isolated from a soil sample collected at Katori, Chiba prefecture, Japan. The strain was cultivated in liquid Waksman medium at 28 °C for 5 days with shaking at 200 rpm. The resulting crude extract was subjected to flash column chromatography using a chloroform/methanol gradient and separated into five fractions. Further purification of these fractions on Sephadex LH-20 and preparative TLC yielded one new natural product (**18**) along with the four known compounds **19**–**22** (Fig. 36) (*66*).
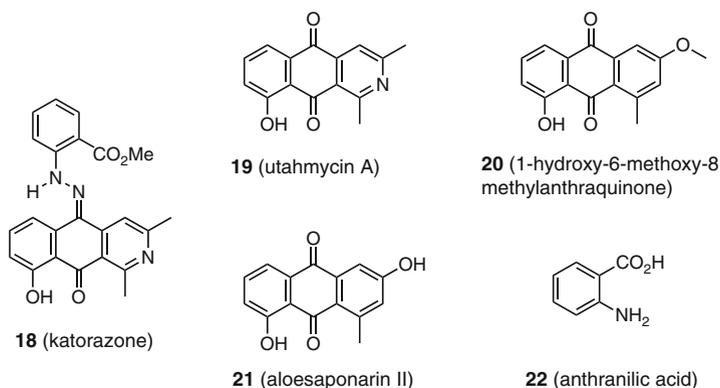
Fig. 36 Katorazone (**18**) and known compounds **19**–**22** isolated from *Streptomyces* sp. IFM 11299

Compound **18**, named katorazone, was isolated as a yellow solid, and its molecular formula was established as $C_{23}H_{19}N_3O_4$ from the HRESIMS data at *m/z* 400.1285 (M-H)⁻ [Δ −1.2 mmu]. The UV spectrum of **18** (λ_{max} 440, 323, and 220 nm) indicated the presence of a quinone chromophore, and the IR absorption bands at 3,558, 1,678, and 1,585 cm⁻¹ implied the presence of hydroxy, carbonyl, and imine groups. Detailed analysis of the 2D-NMR spectra (¹H–¹H COSY, HMQC, and HMBC) of **18** led to the assignments of two units, A and B (Fig. 37a). The low-field resonance
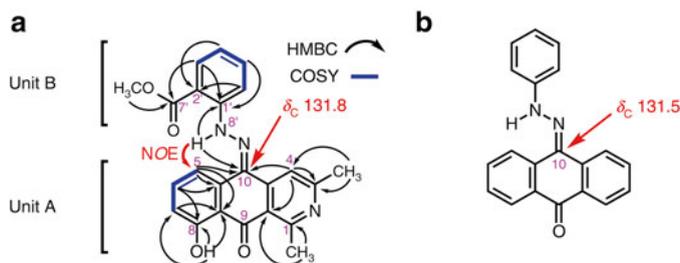
Fig. 37 (**a**) Key COSY and HMBC correlations and ^{13}C NMR chemical shift at C-10 of katorazone (**18**). (**b**) ^{13}C NMR chemical shift at C-10 of 9,10-anthraquinone-9-phenylhydrazone (a synthetic compound)

of the OH-8 signal (δ_H 13.35) suggested an intramolecular hydrogen bond between the hydroxy group proton (OH-8) and the oxygen of the C-9 carbonyl group. On the basis of the predicted molecular formula, a nitrogen atom was incorporated into the right half of unit A. This substructure was supported by the comparison of ^{1}H NMR spectroscopic data with those of utahmycin A (**19**) (*67*). Considering the HMBC correlation of NH-8′ (δ_H 12.72) to C-1′ and C-10, the presence of an imine moiety was deduced by the IR absorption at 1,585 cm^{-1}, and the remaining two unassigned nitrogen atoms, a hydrazine moiety, was proposed and attached to C-1′ of the aryl ring, thereby completing the structure of a phenylhydrazone unit. The low-field-shifted proton (12.72 ppm (1H, s)) suggested an intramolecular hydrogen bond between the NH-8′ and the oxygen of the C-7′ carbonyl group, which supported the presence of *o*-hydrazinylbenzoic acid moiety. The connectivity of the two partial structures A and B was assigned on the basis of HMBC correlations of H-4, H-5, and NH-8′ to C-10 (δ_C 131.8). The chemical shift of C-10 (δ_C 131.8) was consistent with that of the corresponding carbon of synthetic 9,10-anthraquinone-9-phenylhydrazone (δ_C 131.5) (Fig. 37b) (*68*). The NOE observed between NH-8′ (δ_H 12.72) and H-5 (δ_H 7.87) led to the configuration for the diastereomeric C=N bond as shown in Fig. 37a. To the best of our knowledge, katorazone (**18**) is the first example of a 1,3-dimethyl-2-azaanthraquinone-phenylhydrazone to have been isolated from Nature.

6.2 Utahmycin and Others

Four known compounds isolated from the extract of the strain of S*treptomyces* sp. IFM 11299 were identified as utahmycin A (**19**) (*67*), 1-hydroxy-6-methoxy-8-methylanthraquinone (**20**) (*69*), aloesaponarin II (**21**) (*69*), and anthranilic acid (**22**) (*70*) by comparing their spectroscopic data with values reported in the literature (Fig. 36). Utahmycin A (**19**) was produced previously by genetically modified *Streptomyces albus* J1704 transformed with an environmental DNA-derived Erd

gene cluster. To the best of our knowledge, the isolation of utahmycin A (**19**) from the extract of S*treptomyces* sp. IFM 11299 (*66*) was the first report of the isolation of **19** from a wild-type strain. A few 2-azaanthraquinones have been isolated from different fungi (*59–61*). However, farylhydrazones A and B are the only known natural phenylhydrazones produced by *Isaria farinosa* (*71*). Farylhydrazones are phenylhydrazones of pyruvic acid derivatives (Fig. 38).

farylhydrazone A   farylhydrazone B
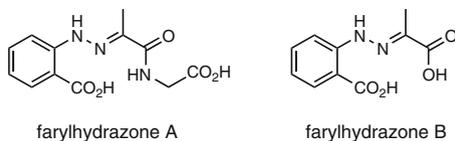
Fig. 38 Farylhydrazones, known natural phenylhydrazones

Katorazone (**18**) could arise biosynthetically from condensation of the carbonyl group at C-10 of utahmycin A (**19**) and methyl 2-hydrazinylbenzoate, which may be derived from anthranilic acid (**22**) (Fig. 39). Two precursors (**19** and **22**) were both isolated from the extract of S*treptomyces* sp. IFM 11299 (*66*).

19 (utahmycin A)

22 (anthranilic acid)   methyl 2-hydrazinylbenzoate
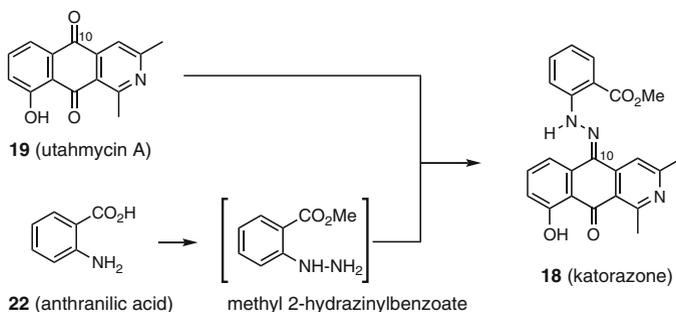
18 (katorazone)

Fig. 39 A proposed biogenetic scheme of katorazone (**18**) from utahmycin A (**19**) and anthranilic acid (**22**)

6.3   Effects on Wnt and TRAIL Signaling

Katorazone (**18**) was evaluated for Wnt signal inhibition activity and TRAIL-resistance overcoming activity. The assay results (Fig. 40 top) showed that katorazone (**18**) showed moderate inhibition of Wnt signal transcriptional activity with IC_{50} values of *ca.* 50 μ*M*.

The bioactivity of katorazone (**18**) was also evaluated for effects on TRAIL resistance in AGS (TRAIL-resistant human gastric adenocarcinoma) cells, by comparing cell viability in the presence and absence of TRAIL (100 ng/cm^3) using the
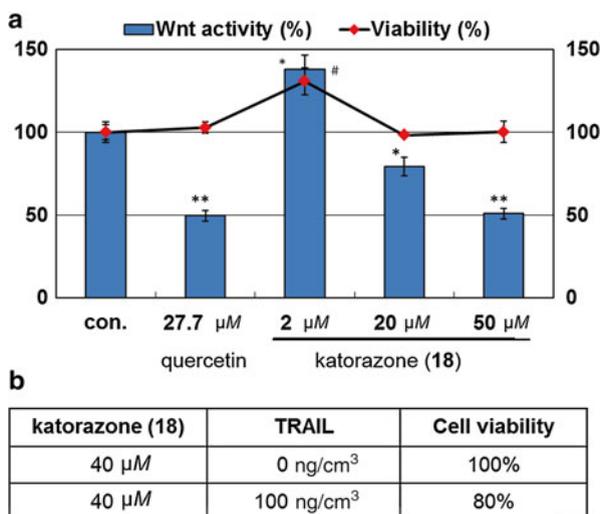
Fig. 40 Wnt signal (*top*) inhibitory activity of katorazone (**18**). Bar chart showing Wnt signal transcriptional activity and a line chart representing viability of the STF/293 cells. Quercetin was used as a positive control. TRAIL-resistance overcoming activity tests in AGS (TRAIL-resistant human gastric adenocarcinoma) cells (*bottom*)

fluorometric microculture cytotoxicity assay (FMCA) (Sect. 3.2, Fig. 6). With 40 μM of **18**, the cell viability was 100%, whereas the same concentration of **18** in the presence of 100 ng/cm^3 TRAIL reduced cell viability to 80% of control levels (Fig. 40 bottom). These results suggest that katorazone (**18**) showed a moderate synergistic effect in combination with TRAIL using AGS cells.

7 Fuzanins

7.1 *Fuzanins A–D*

Field samples for the isolation of actinomycete strains were collected not only in Chiba area but also in various other locations in Japan. Culture extracts of more than 300 actinomycete strains were subjected to preliminary screening for antimicrobial activity as well as TLC examination using anisaldehyde as a spray reagent. A strain of *Kitasatospora* sp. IFM10917 (Fig. 41), which was separated on humic acid-vitamin (HV) agar medium from a soil sample collected in March 2007 at Toyama Castle Park, Toyama, Japan, was selected for investigation of antimicrobial activity using a chemical screening method.

The culture supernatant of *Kitasatospora* sp. IFM10917 was partitioned between EtOAc and 10% aqueous MeOH, and the aqueous phase was further extracted with *n*-BuOH to give three fractions. The EtOAc-soluble fraction was subjected to

Fig. 41 *Kitasatospora* sp. IFM 10917
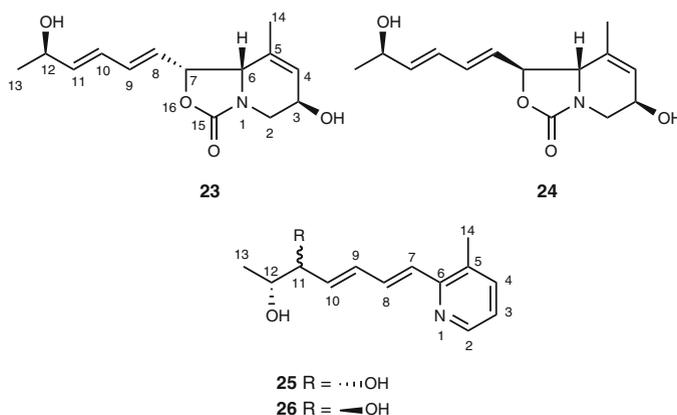
Fig. 42 Fuzanins A-D (**23–26**)

passage over columns containing Sephadex LH-20 and silica gel, followed by final purification using reversed-phase HPLC on ODS, to give four new compounds (**23–26**) (Fig. 42) (72). These new compounds were named fuzanins A (**23**), B (**24**), C (**25**), and D (**26**), because 'Toyama' in Chinese characters can be also read as 'fuzan' in Japanese.

Fuzanin A (**23**) gave a molecular formula of $C_{14}H_{19}NO_4$, suggested from its HRFABMS data (*m/z* 304.0976 [M+K]$^+$, Δ +2.4 mmu), and this formula indicated six degrees of unsaturation, of which four were accounted for by one carbonyl group and three double bonds (δ_C 125.4, 126.8, 129.5, 131.1, 135.5, and 142.4 ppm). Hence, the remaining two degrees of unsaturation indicated the presence of two rings. The ^{13}C NMR chemical shift of the carbonyl group (C-15, δ_C 155.9 ppm)

indicated it to be located between an oxygen and a nitrogen atom, suggesting the presence of a carbamate moiety. Analysis of the COSY spectrum of **23** was used to reveal two partial structures, a C-2 to C-4 residue and a C-6 to C-13 chain containing one diene moiety (C-8 to C-11) and one methyl group (C-14). The large coupling constants ($J_{8,9} = 15.2$ Hz and $J_{10,11} = 15.0$ Hz) both indicated *trans*- configurations. The HMBC spectrum showed the ^{1}H–^{13}C long-range correlations for H$_2$-2/C-15, H-7/C-15, and H$_2$-2/C-6, indicating the presence of five-membered ring to construct a carbamate group (N-1, C-6, C-7, O-16, and C-15). The presence of a 3-methyltetrahydropyridine ring was deduced by HMBC correlations for H$_3$-14/C-4, H$_3$-14/C-5, H$_3$-14/C-6, and H-4/C-2. The N*O*E correlations observed for H-2α/H-3, H-2β/H-6, and H-6/H-7 revealed that H-2β, H-6, and H-7 are *cis*, and H-2α and H-3 are also *cis*. This was also supported from the large *vicinal* coupling constant ($J = 11.4$ Hz) between H-2β and H-3, indicating that H-2β and H-3 have a *trans-diaxial* relationship. The absolute configurations at C-3 and C-12 were elucidated by comparing the ^{1}H NMR chemical shifts of the (*S*)- and (*R*)- MTPA diesters of **23**, on the basis of the modified *Mosher*'s method, to lead to (3*R*,6*S*,7*R*)- and (12*R*)-configurations.

Fuzanin B (**24**) was assigned the same molecular formula, $C_{14}H_{19}NO_4$, as fuzanin A (**23**). Analysis of the COSY and HMBC spectroscopic data revealed that the planar structure of **24** is identical to that of **23**. The ^{1}H and ^{13}C NMR chemical shifts in DMSO-d_6 of **24** were similar to those of **23**, except for the signals at the C-6, C-7, and C-8 positions. The ^{1}H NMR signals of H-6, H-7, and H-8 for **23/24** resonated at δ_H 4.38/3.98 ppm, δ_H 5.17/4.66 ppm, and δ_H 5.46/5.81 ppm, respectively, and the ^{13}C NMR signals of C-6, C-7, and C-8 for **23/24** resonated, in turn, at δ_C 58.0/60.3 ppm, δ_C 78.3/79.9 ppm, and δ_C 125.4/128.8 ppm. From these observations, fuzanin B (**24**) was proposed as a stereoisomer of fuzanin A (**23**) at the C-7 position, consistent with the significant N*O*E observation between H-6 and H-8 for **24** as opposed to the N*O*E correlation apparent between H-6 and H-7 for **23**.

The molecular formula of fuzanin C (**25**) was determined as $C_{13}H_{17}NO_2$ on the basis of high-resolution EIMS (*m/z* 219.1293 [M$^+$], Δ +3.4 mmu). The planar structure of **25** was elucidated by analysis of the HMBC, HMQC, and COSY spectra to indicate the presence of a disubstituted pyridine ring, possessing a side chain at C-6 with a conjugated diene and a *vicinal* diol. The relatively large vicinal coupling constant ($J_{11,12} = 6.5$ Hz) of the 1,2-diol moiety in **25** revealed a *syn* relationship between H-11 and H-12, since *anti* and *syn* 1,2-diols in the literature having corresponding partial structures were described with the coupling constants 3.0 Hz when *anti* (*73*), and 6.2 Hz when *syn* (*74*). The absolute configuration of the *syn*-1,2-diol moiety was deduced by the CD spectrum in the presence of Ni(acac)$_2$, which showed a positive *Cotton* effect at 313 nm and negative *Cotton* effect at 286 nm, indicating (11*R*) and (12*R*) configurations, according to Ref. (*75*).

Fuzanin D (**26**) was revealed to have the same planar structure as fuzanin C (**25**), and **26** had a relatively small *vicinal* coupling constant ($J_{11,12} = 3.8$ Hz) for the 1,2-diol moiety, suggesting the *anti*-relationship between H-11 and H-12, on the basis of comparison with a literature value (*73*).

Fuzanins A–D (**23–26**) were isolated and characterized as unique natural products containing a carbamate or pyridine moiety from an actinomycete. No compounds having related structures are known from *Kitasatospora* species, and the biogenesis of fuzanins is yet to be elucidated. Fuzanin D (**26**) exhibited weak cytotoxicity against DLD-1 human colon carcinoma cells, with an IC_{50} value of 41.2 μ*M*, while fuzanins A-C (**23–25**) proved to be inactive (IC_{50} values of >50 μ*M*). Fuzanin D (**26**) also exhibited moderate inhibition of Wnt signal transcription (inhibition: 62%) along with weak cytotoxicity (28%) at 25 μ*M* when it was tested for its Wnt signal inhibitory activity using a luciferase reporter gene assay in SuperTOP-Flash transfected cells (Sect. 3.3.3).

7.2   Fuzanins E-I

Fuzanins A-D (**23–26**) are new natural products containing a carbamate or pyridine moiety, and a hint of their biogenesis may be provided by further detailed analysis of secondary metabolites including their precursors from this strain. Thus, our group further investigated the *Kitasatospora* sp. IFM10917 culture and performed chemical studies on its crude extract, resulting in the isolation of additional carbamate-containing metabolites, named fuzanins E-I (**27–31**) (Fig. 43) (*76*).
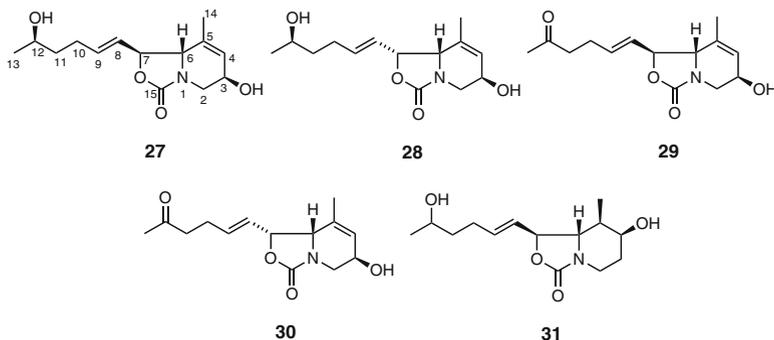
Fig. 43   Fuzanins E-I (**27–31**)

The structure elucidation of fuzanins E-I (**27–31**) was carried out mainly based on comparison of their spectroscopic data with those of fuzanins A (**23**) and B (**24**), to reveal that the structural differences were evident in their side chains. Fuzanins A (**23**) and B (**24**) have a diene moiety at the C-8 through C-11 positions, while fuzanins E- I (**27–31**) all possess a monoene at the C-8/C-9 position. In addition, fuzanins E (**27**), F (**28**), and I (**31**) have a secondary hydroxy group at C-12, whereas fuzanins G (**29**) and H (**30**) possess a carbonyl group at C-12. The relative configurations of H-6 and H-7 of **27–31** were assignable by comparison of the chemical shifts of H-6, H-7, and H-8 (Fig. 44).
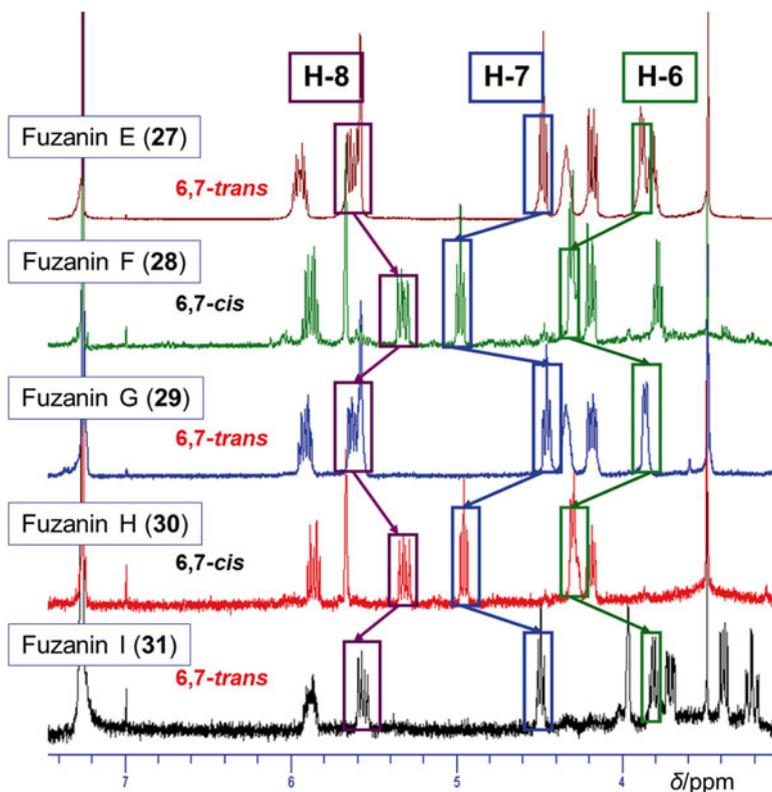
Fig. 44 Comparison of the ¹H NMR chemical shifts of H-6, H-7, and H-8 of fuzanins E-I (**27–31**) in CDCl₃ and assignment of the H-6 and H-7 configurations

The bioactivities of fuzanins E-I (**27–31**) were determined using a cell-based assay system targeting the Wnt signaling pathways, but these compounds did not exhibit inhibition of Wnt signal transcription activity even at 100 μ*M* using a luciferase reporter gene assay in SuperTOP-Flash transfected cells. Fuzanins E-I (**27–31**) also did not show antimicrobial activity at 100 μg/disk against *Bacillus subtilis*.

8 Elmonin

Based on a preliminary screening study for the selection of interesting actinomycete strains, involving cytotoxic activity testing against human gastric adenocarcinoma (AGS) cells, as described in Sect. 3.2, and, after examining extracts of 559 strains of actinomycetes, a strain of *Streptomyces* sp. IFM11490 was selected as being worthy of further investigation. This strain was separated from a soil sample collected at the main gate of Hokkaido University, Sapporo, Japan, and the cultured broth (12 L
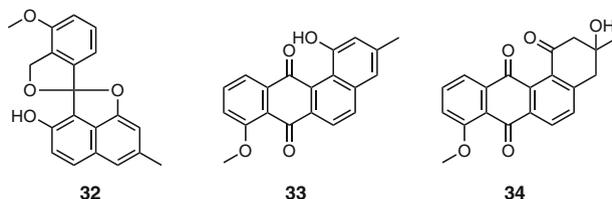
Fig. 45 Elmonin (**32**) and the two known compounds **33** and **34** isolated from *Streptomyces* sp. IFM11490

Waksman medium, in total) was harvested and centrifuged to separate the mycelium and supernatant. The supernatant was extracted with ethyl acetate three times to give 3.6 g of extract, which was subjected to silica gel and Sephadex LH-20 column chromatography to give three compounds **32–34** (Fig. 45). Two of these were identified as known substances, X14881E (**33**) (*77*) and 6-deoxy-8-methylrabelomycin (**34**) (*77*), by comparison of spectroscopic data with literature values. Compound **32** proved to be a new compound, and its structure was elucidated on the basis of analysis of its spectroscopic data. A patent application concerning this new compound, elmonin (**32**), was filed in 2013 (*78*), while a compound with an identical planar structure was published later in 2013 by *R. Müller* and his group as a new compound, named oleaceran (*79*).

Compounds **32–34** from *Streptomyces* sp. IFM11490 exhibited moderate cytotoxicity against AGS cells with IC_{50} values of *ca.* 50, 12.4, and 37.4 μM.

9 Actinomycete Metabolites Found in Screening Studies Targeting Cancer-Related Signaling Pathways

9.1 Teleocidin with Enhancing Death-Receptor 5 Promoter Activity

During our screening studies using a luciferase assay to identify natural products that enhance death receptor 5 (DR5) expression, the DR5 promoter activity of 656 extracts of actinomycetes were examined. As a result, an extract of CKK609 strain was selected as an active sample (4.0-fold increase in DR5 promoter activity at 500 $\mu g/cm^3$). This strain was isolated from a soil sample collected at Izumi Forest, Chiba, Japan, in June 2007. The spores grown on solid Waksman medium were inoculated into a 500 cm^3 *Sakaguchi* flask containing 100 cm^3 of liquid medium and cultured for 4 days at 28 °C with reciprocal shaking at 200 rpm to produce a seed culture. A 10 cm^3 aliquot of the seed culture was transferred into a 3 dm^3 flask containing 500 cm^3 of the same medium and cultured for 5 days at 28 °C with reciprocal shaking at 100 rpm. The fermentation broth (32 dm^3) was centrifuged and the
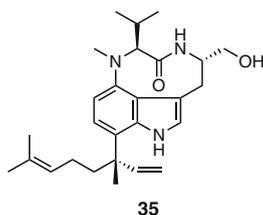
Fig. 46 Teleocidin A-2 (**35**)

mycelium was extracted with MeOH for one day. When the extracts were filtered, concentrated *in vacuo*, and partitioned with EtOAc, the DR5 promoter activity was concentrated in the EtOAc-soluble fraction (4.5-fold at 50 μg/cm³). Fractionation of the EtOAc fraction by column chromatography, followed by reversed-phase HPLC, afforded teleocidin A-2 (**35**, Fig. 46) (*80*), which was previously isolated from *Streptomyces mediocidicus* (*81*).

The isolated teleocidin A-2 (**35**) was tested for DR5 promoter activity using a luciferase assay in DLD-1/*Sac*I cells, and luteolin was used as a positive control (*45*) at 17.5 μ*M* to reveal that teleocidin A-2 (**35**) increased DR5 promoter activity in a dose-dependent manner as compared with control cells (3.9-fold at 0.01 μ*M* and 4.7-fold at 0.1 μ*M*).

Teleocidin A-2 (**35**) was next examined for its ability to overcome TRAIL resistance in AGS cells. AGS human gastrointestinal tract cancer cells were reported recently to be refractory to apoptosis induction by TRAIL (*82*). To assess the possible combination effect of teleocidin A-2 (**35**) and TRAIL on cell viability, AGS cells were treated with teleocidin A-2 (**35**) alone, TRAIL alone (200 ng/cm³), and a combination treatment with TRAIL for 24 h, and subjected to an FMCA evaluation procedure (*9*). However, combined treatment with teleocidin A-2 (**35**) and TRAIL did not show any change in cell viability, when compared to the TRAIL treatment alone. To clarify the underlying mechanism of enhancement of DR5 promoter activity by teleocidin A-2 (**35**), further studies are needed.

9.2 Tyrosine Derivatives with TRAIL-Resistance Overcoming Activity

In our screening program for the isolation of natural products that abrogate TRAIL resistance, MeOH extracts of 120 culture broths of actinomycete strains collected from different areas of Japan were examined by comparing cell viability in the presence and absence of TRAIL against a TRAIL-resistant gastric adenocarcinoma (AGS) cell line. Among 23 active strains, strain *Streptomyces* sp. IFM 10973, isolated from a soil sample collected at Sakazuki Forest, Chiba, Japan, was selected for bioassay-guided fractionation, since a concentration of 1,000 μg/cm³ of this extract

showed more than a 30% difference in cell viability against AGS cells in the presence and absence of TRAIL (100 ng/cm^3). The fermentation broth (5 dm^3) of *Streptomyces* sp. IFM 10973 was centrifuged, and the supernatant was partitioned between EtOAc and *n*-BuOH. The *n*-BuOH extract, which produced 25% more inhibition of cell viability against AGS cells than the TRAIL alone at 200 μg/cm^3, was subjected to ODS and silica gel column chromatographic separations, followed by ODS HPLC, to yield two new prenyltyrosine derivatives (**36** and **37**) (Fig. 47) (*83*). The EtOAc extract, which produced 30% more inhibition at 50 μg/cm^3, was subjected to ODS flash column chromatography to afford a known compound, novobiocin (**38**) (*84*).

36 R = COCH$_3$
37 R = H

38
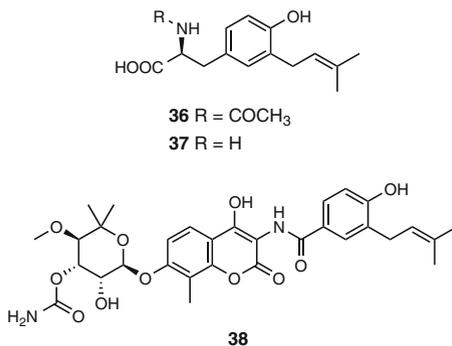
Fig. 47 Two new prenyltyrosine derivatives **36** and **37** and novobiocin (**38**)

Structures of two new compounds (**36** and **37**) were elucidated from their spectroscopic data to suggest the presence of 1,2,4-trisubstituted benzene ring and a 3-methyl-2-butenyl (prenyl) group. Analysis of its COSY and HMBC spectra led to the planar structure of **36** as a prenyltyrosine derivative with an acetamide group on C-8 (Fig. 48). Compound **37** possesses a free amino group at C-8. The absolute configuration of the C-8 position was determined as (*S*) (L-tyrosine derivative) by a modification of *Marfey*'s method (*85*) using D-FDLA (1-fluoro-2,4-dinitrophenyl-5-D-leucinamide) and L-FDLA as derivatizing agents (*86*).
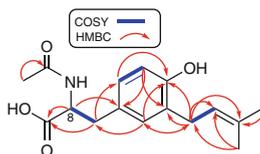
Fig. 48 Key COSY and HMBC correlations of **36**

The isolated compounds **36**–**38** from *Streptomyces* sp. IFM 10973 were evaluated for their ability to overcome TRAIL-resistance in AGS cells. AGS cells were treated with each of these compounds (**36**–**38**) alone, TRAIL alone, and a combination treatment with TRAIL, and then subjected to a FMCA procedure (*9*). As shown in Fig. 49, treatment with 100 ng/cm^3 TRAIL for 24 h resulted in only a slight decrease in cell viability (92±4%), which was similar to the effects of **36** and **38**
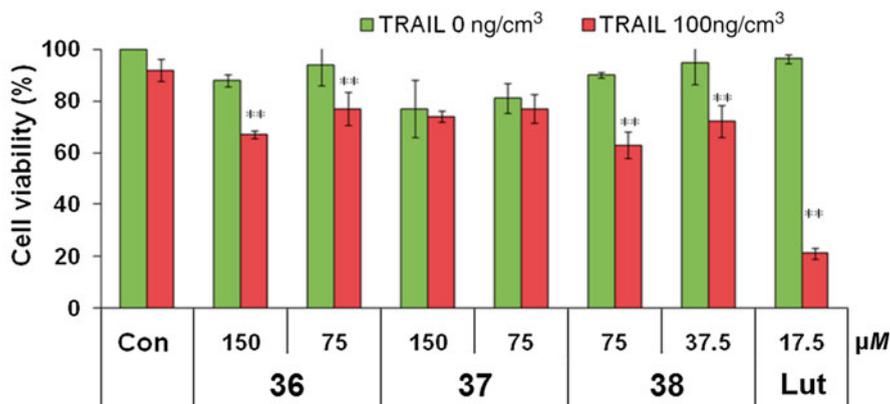
Fig. 49 Effect of the isolated compounds **36–38**, luteolin (positive control: Lut), and TRAIL treatment, alone and in combination, on the viability of AGS cells

alone. Luteolin, used as the positive control (*45*), produced about 70% greater inhibition along with TRAIL than the TRAIL alone. Treating cells with 100 ng/cm³ TRAIL and 75 or 150 μ*M* of **36** reduced cell viability to $77 \pm 7\%$ and $67 \pm 5\%$ of control levels ($p < 0.01$), respectively, which was 15 and 25% more than the TRAIL alone, suggesting a possible synergism between the two agents. Combined treatment with TRAIL and **37** did not produce any significant reduction in the cell viability suggesting its inactivity in overcoming TRAIL resistance. On the other hand, treatment with TRAIL and 37.5 or 75 μ*M* of **38** reduced cell viability to $72 \pm 6\%$ and $63 \pm 5\%$ of controls ($p < 0.01$), respectively, which were 20 and 29% more than the TRAIL alone. These results may suggest its possible synergistic activity in combination with TRAIL against AGS cells.

Resistance of cancer cells toward TRAIL may occur at different points in the TRAIL-induced apoptotic pathways. Understanding the mechanisms of such resistance and developing strategies to overcome it are important for the successful use of TRAIL in cancer therapy. Combined treatment of TRAIL and chemotherapeutic agents including some natural products can overcome such resistance, and sensitize TRAIL-resistant cells to enhance the therapeutic potential of TRAIL against the cancer cells. Therefore, a natural product producing synergistic activity with TRAIL would be a new tool for cancer therapy (*10*, *11*).

9.3 Nonactins, Griseoviridin, and Nocardamines with Wnt Signaling Inhibitory Activity

During a search for bioactive natural products targeting Wnt signaling, and a series of extracts of cultured actinomycetes using a cell-based luciferase assay was examined to evaluate the inhibition of TCF/β-catenin transcriptional activity (TOP activity), which was measured using the STF/293 cell line (a 293 human embryonic

kidney cell line stably transfected with SuperTOP-Flash) (Sect. 3.3.3). As a result, it was found that an extract of strain CKK179 showed an 88% inhibition with high (>85%) cell viability at 500 μg/cm^3. This strain was isolated in 2005 from a soil sample collected at Sakazuki Forest, Chiba, Japan, and fractionation of this active strain was carried out (87). The fermentation broth of CKK179 was centrifuged and the supernatant was extracted with EtOAc. The EtOAc-soluble fraction was subjected to chromatography on a silica gel column followed by passage over a Sephadex LH-20 column to give nonactin (39), monactin (40), and dinactin (41) (Fig. 50) (88), which were identified by ESIMS as well as HPLC using commercially available standards. The inhibitory effects of Wnt signal activity by these three compounds 39–41 were compared with a related compound, dimeric dinactin (42), which was isolated from another *Streptomyces* sp. (YM09-028) by *S. Tsukamoto*'s group at Kumamoto University (87).

Wnt signaling activates gene transcription by forming a complex between DNA-binding proteins of the TCF/LEF family and β-catenin. SuperTOP-Flash, a β-catenin-responsive reporter plasmid with multiple TCF-binding sites (CCTTTGATC), was activated in cells. SuperFOP-Flash has eight mutated TCF-binding sites (CCTTTGGCC), and a selective inhibitor would prevent any enhancement of transcription in SuperFOP-flash-transfected cells; thus the ratio of TOP/FOP-flash reporter activity provides a measure of selective inhibition of Wnt signaling.

The inhibition of TCF/β-catenin transcriptional activity (TOP-Flash activity) by 39–42 was examined along with effects on cell viability since a decrease in cell
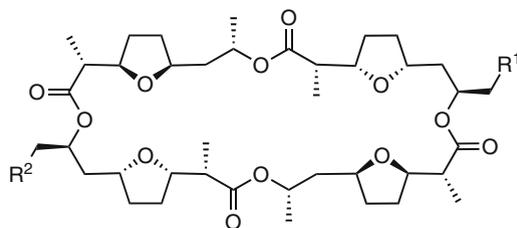
39 R^1 = R^2 = H

40 R^1 = CH$_3$, R^2 = H

41 R^1 = R^2 = CH$_3$
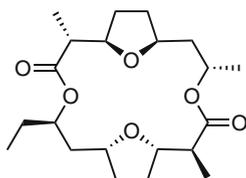
42

Fig. 50 Nonactin (39), monactin (40), dinactin (41), and dimeric dinactin (42)
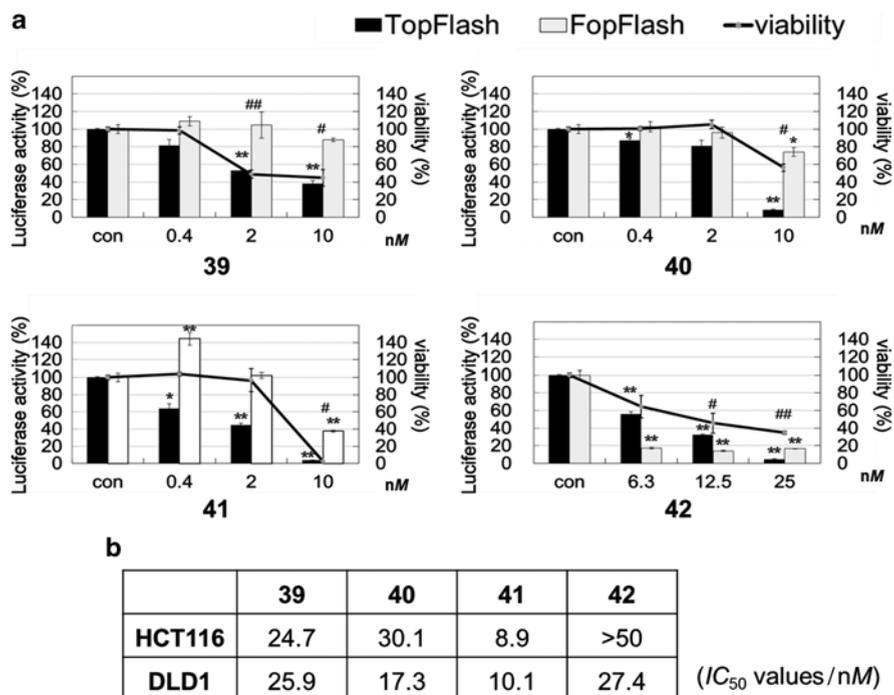
Fig. 51 (**a**) Effects of **39**–**42** on TOP- and FOP-flash activity and cell viability. Quercetin was used as positive control and showed 29% of TOP-Flash activity and 94% of viability at 27.7 μM. (**b**) Cytotoxicity of **39**–**42** against selected colon cancer cell lines (IC_{50} values/nM)

number may contribute to the inhibition (Fig. 51a). These compounds inhibited TOP-Flash activity with IC_{50} values of 7.2, 5.0, 1.3, and 7.4 nM, respectively. The FOP-Flash activity was also examined, and compounds **39**–**41** showed weak inhibition of FOP-flash activity at the same concentration as their IC_{50} values, while compound **42** caused a greater decrease in FOP-Flash activity and cell viability, implying that it did not selectively inhibit Wnt signaling. Considering the cell viability data (Fig. 51a), compounds **39** and **40** showed 20% inhibition of TOP-Flash activity at 0.4 nM and 2 nM, while compound **41** showed 60% inhibition of TOP-Flash activity at 2 nM. Thus, compound **41** showed the most potent inhibition of TOP-Flash activity. The inhibition of TCF/β-catenin transcriptional activity may be caused by β-catenin degradation or inhibition of β-catenin translocation into the nucleus. Further investigation of the molecular mechanism of these compounds is warranted. The cytotoxicity of these compounds for two colon cancer cell lines, HCT116 and DLD1, was examined, since the Wnt signal pathways of many human colorectal cancer cells are known to be activated aberrantly. Compounds **39**–**42** were found to exhibit cytotoxicity with the IC_{50} values shown in Fig. 51b. Nonactin is known as an ionophore of ammonium and potassium ions and its effect on Wnt signaling

pathway has not been reported previously, while the calcium ionophore A23187 was documented to inhibit the Wnt/β-catenin pathway through activation of protein kinase C, and promotes the phosphorylation of β-catenin, resulting in proteasome-mediated β-catenin degradation (*89*).

On further examination of extracts of actinomycete samples (*90*), it was found that the methanol extracts of fermented broth and mycelia of strains CKK748 and CKK784 obtained on a small scale exhibited marked activity (98% and 96% inhibition at 10 μg/cm^3 with high (>90%) cell viability, respectively) based on a cell-based luciferase assay system to evaluate the inhibition of the transcriptional activity of TCF/β-catenin (TOP-flash activity) with the STF/293 cell line. Strain CKK748 was isolated from a sample of sea water collected in Kujukuri-Hama, Chiba, Japan, in October 2009, while the strain CKK784 was separated from a soil sample collected in Chiba, Japan in August 2010. Activity-guided fractionation of the culture broth and mycelium of actinomycete CKK748 led to the isolation of griseoviridin (**43**) (*91*). From the culture broth and mycelium of actinomycete CKK784, activity-guided fractionation was carried out by silica gel and ODS column chromatography, followed by reversed-phase HPLC. This afforded four cyclic hydroxamate compounds, nocardamine (**44**) (*92*), dehydroxynocardamine (**45**) (*92*), desmethylenylnocardamine (**46**) (*92*), and bisucaberine (**47**) (*93*), as the active components in the supernatant, together with *N*-formylantimycic acid methyl ester (**48**) (*94*) from the mycelium (Fig. 52).

Compounds **43**–**45** exhibited inhibition of SuperTOP-flash activity with IC_{50} values of 13, 14, and 8.2 μ*M*, with high viability (>80%), and a weak or moderate decrease in SuperFOP-flash activity, while **46** did not show any dose-dependent inhibitory activity (Fig. 53). Compound **47** decreased the SuperTOP-flash activity
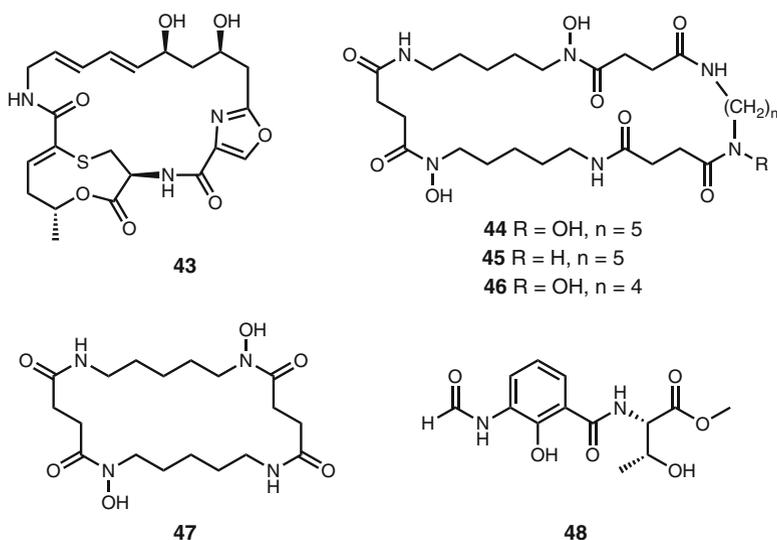
Fig. 52 Griseoviridin (**43**), nocardamine (**44**), dehydroxynocardamine (**45**), desmethylenylnocardamine (**46**), bisucaberine (**47**), and *N*-formylantimycic acid methyl ester (**48**)
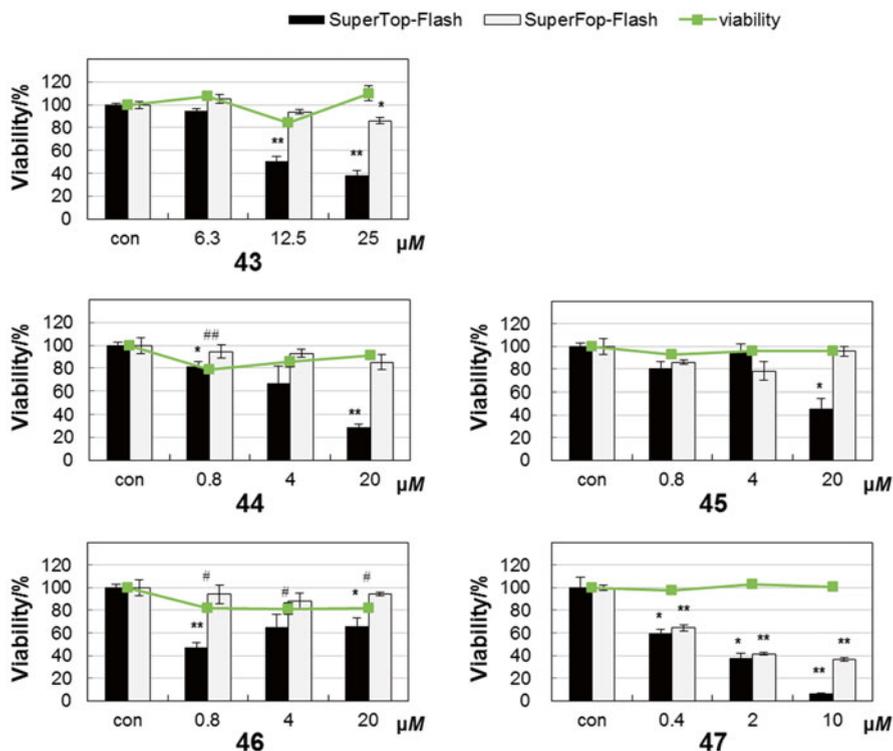
Fig. 53 Effects of griseoviridin (**43**) and cyclic hydroxamate compounds **44**–**47** on SuperTOP-Flash and SuperFOP-Flash activity in STF/293 and 293 T cells determined by a luciferase assay, and viability in STF/293 cells

with an IC_{50} of 1.1 μ*M*. However, **47** (0.4–2 μ*M*) also decreased SuperFOP-flash activity, implying that it did not inhibit Wnt signaling selectively at 0.4–2 μ*M*. Compound **48** (2–50 μ*M*) did not show any inhibition of SuperTOP-Flash activity. The cytotoxicity of these compounds (**43**–**47**) was examined against DLD1, HCT116, and SW480 human colorectal cancer cells, in which the Wnt signaling pathway is known to be activated. Compound **43** showed 48–62% viability against DLD1 cells, *versus* more than 60% viability against HCT 116 and SW480 cells at 6.3–50 μ*M*. From the results obtained, compound **43** was weakly cytotoxic to DLD1 cells, but not to HCT 116, or SW480 cells, while compounds **44**–**46** did not show cytotoxicity against DLD1, HCT116, or SW480 cells (the viability was more than 60% at 0.8–20 μ*M*). Although these compounds (**43** and **44**–**47**) inhibited TCF/β-catenin transcriptional activity, they did not show significant cytotoxicity against the colorectal cancer cell lines in which they were evaluated. While griseoviridin (**43**) and the cyclic hydroxamate compounds (**44**–**47**) are known as a broad-spectrum antibiotic and siderophores, respectively, their effect on Wnt signaling has not been reported previously. Compounds **43**–**45** may warrant further study as potential Wnt signaling inhibitors.

10 Synthesis Aspects

10.1 Synthesis of Phenazines

This section deals with synthesis aspects of heterocyclic natural products described
in this section studied by other groups, particularly on phenazines (Sect. 10.1) and
pyranonaphthoquinones (Sect. 10.2).

Synthesis studies of phenazine alkaloids were previously summarized in a review
in 2004 (*31*). Subsequently, a facile electrochemical synthesis of phenazine deriva-
tives was reported. The electrooxidation of 2,3-dimethylhydroquinone (**49**) in the
presence of *o*-phenylenediamines (**50**) afforded the corresponding phenazine deriv-
ative **51** in a good yield (Fig. 54) (*95*).
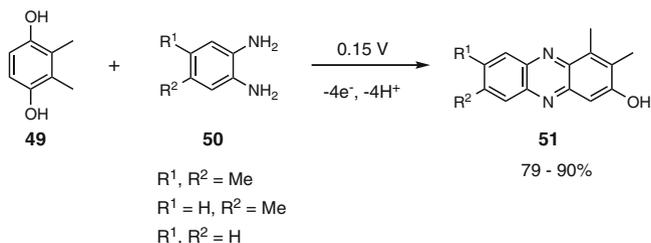
Fig. 54 Electrochemical synthesis of phenazine derivatives

Recently, a new method for biomimetic synthesis of phenazine-1,6-dicarboxylic
acid (PDC) was reported (Fig. 55) (*96*). 1,2-*trans*-2-Amino-3-oxocyclohexane-1-
carboxylate (**52**), resembling the proposed biosynthetic substrate, underwent facile
dimerization and oxidation in air to a tetrasubstituted pyrazine (**54**). Oxidation and
saponification delivered the natural product PDC (**5**, Sect. 4.1, Fig. 13).
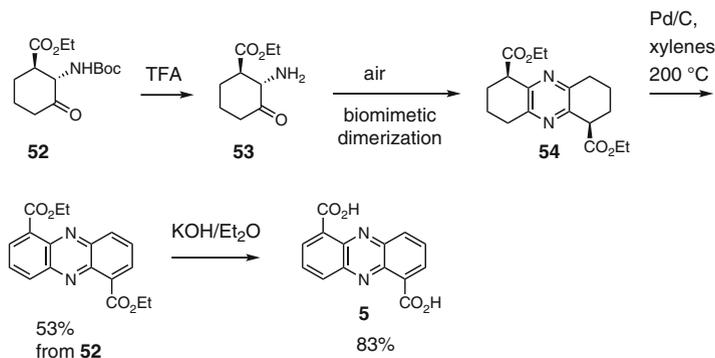
Fig. 55 Biomimetic synthesis of phenazine-1,6-dicarboxylic acid (PDC) (**5**)

10.2 Synthesis of Pyranonaphthoquinones

Synthesis of pyranonaphthoquinone antibiotics was studied extensively in the early 1980s, and determination of absolute configuration of the natural product griseusin A was achieved by total synthesis of its (+)-enantiomer (*97*). Studies on the synthesis of pyranonaphthoquinones have been continued by many groups, with the results obtained summarized (*98*, *99*). Further specific studies on the synthesis of pyranonaphthoquinone derivatives reported recently are described below.

A pyranonaphthoquinone (**55**) was prepared by *Diels-Alder* reaction of pyranoquinone (**56**) with a diene, (4,4-dimethoxybuta-1,3-dien-2-yloxy)trimethylsilane (**57**), under *Jones'* oxidation conditions, in 85% yield, and **55** was converted to the dimeric natural product crisamicin A (**58**) (Fig. 56) (*100*).
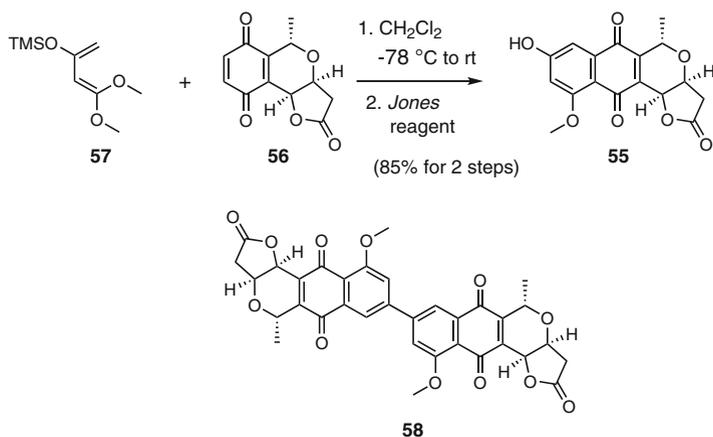
Fig. 56 Key *Diels-Alder* reaction in total synthesis of crisamicin A (**58**)

Synthesis of the core framework (**59**) of the natural product griseusin B (**60**) was achieved by a highly convergent manner. The key step was an efficient one-pot *Hauser-Kraus*-annulation-methylation-double deprotection-spirocyclization sequence that directly afforded the target parent tetracyclic ring system (Fig. 57) (*101*).
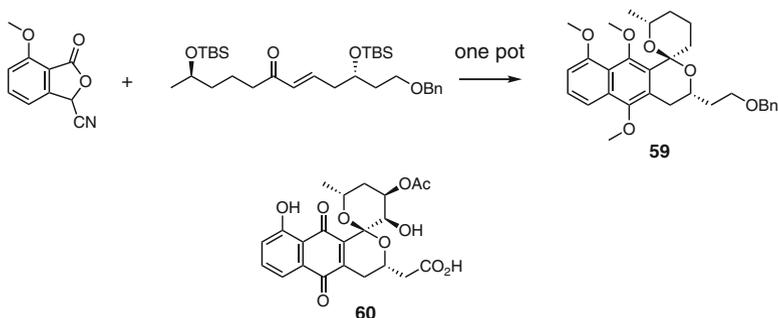
Fig. 57 Key one-pot reaction in the synthesis of the griseusin (**60**) scaffold **59**
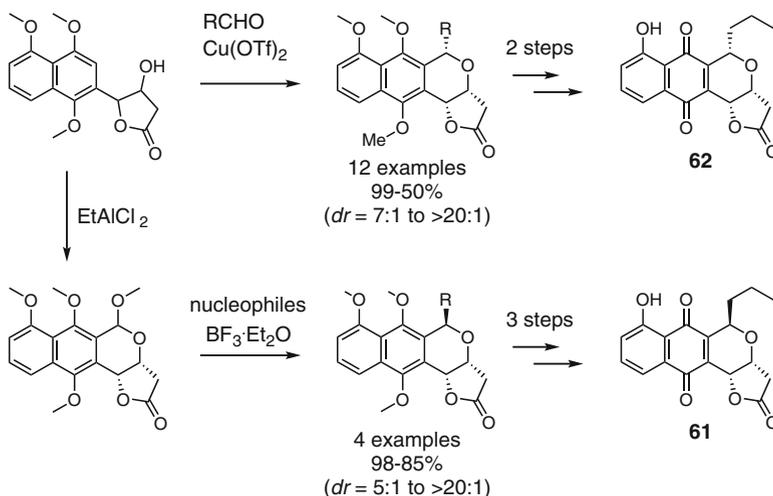
Fig. 58 Key diastereoselective oxa-*Pictet-Spengler* reaction strategy in the total synthesis of (+)-frenolicin B, *epi*-(+)-frenolicin B, and their analogs

Total synthesis of the naturally occurring pyranonaphtoquinones (+)-frenolicin B (**61**) and (+)-*epi*-frenolicin B (**62**), and their precursors, to afford a diverse array of pyranonaphthoquinone analogs, was achieved by an efficient diastereoselective oxa-*Pictet-Spengler* reaction strategy, which was developed to construct benzoisochroman diastereomers (Fig. 58) (*102*).

References

1. Ishibashi M, Arai MA (2009) Search for Bioactive Natural Products Targeting Cancer-Related Signaling Pathways. J Synth Org Chem Jpn 67:1094
2. Ishibashi M, Arai MA (2012) Bioactive Natural Products from Myxomycetes Having Effects on Signaling Pathways. Heterocycles 85:1299
3. Ishibashi M (2007) Study on Myxomycetes as a New Source of Bioactive Natural Products. Yakugaku Zasshi 127:1369
4. Ishibashi M, Yamaguchi Y, Hirano YJ (2006) Bioactive Natural Products from Nudibranchs. In: Fingerman M, Nagabhushanam R (eds) Biomaterials from Aquatic and Terrestrial Organisms. Science Publishers Inc, Enfield, NH, USA, p 513
5. Ishibashi M, Toume K, Yamaguchi Y, Ohtsuki T (2004) Recent Advances in Search for Bioactive Natural Products from Medicinal Plants of North-East Thailand. In: S. G. Pandalai (ed) Recent Research Developments in Phytochemistry, vol 8. Research Signpost, Trivandrum, India, p 139
6. Ahmed F, Sadhu SK, Ishibashi M (2010) Search for Bioactive Natural Products from Medicinal Plants of Bangladesh. J Nat Med 85:393
7. Hayakawa M, Nonomura H (1987) Humic Acid-Vitamin Agar, as a New Medium for the Selective Isolation of Soil Actinomycetes. J Ferment Technol 65:501
8. Hirsch CF, Christensen DL (1983) Novel Method for Selective Isolation of Actinomycetes. Appl Environ Microbiol 46:925
9. Lindhagen E, Nygren P, Larsson R (2008) The Fluorometric Microculture Cytotoxicity Assay. Nat Protoc 3:1364
10. Hellwig CT, Rehm M (2012) TRAIL Signaling and Synergy Mechanisms Used in TRAIL-Based Combination Therapies. Mol Cancer Ther 11:3
11. Amm HM, Oliver PG, Lee CH, Li Y, Buchsbaum DJ (2011) Combined Modality Therapy with TRAIL or Agonistic Death Receptor Antibodies. Cancer Biol Ther 11:431
12. Ishibashi M, Ohtsuki T (2008) Studies on Search for Bioactive Natural Products Targeting TRAIL Signaling Leading to Tumor Cell Apoptosis. Med Res Rev 28:688
13. Kikuchi H, Ohtsuki T, Koyano T, Kowithayakorn T, Sakai T, Ishibashi M (2007) Brandisianins A-F, Isoflavonoids Isolated from *Millettia brandisiana* in a Screening Program for Death-Receptor Expression Enhancement Activity. J Nat Prod 70:1910
14. Kikuchi H, Ohtsuki T, Koyano T, Kowithayakorn T, Sakai T, Ishibashi M (2009) Death Receptor 5 Targeting Activity-Guided Isolation of Isoflavones from *Millettia brandisiana* and *Ardisia colorata* and Evaluation of Ability to Induce TRAIL-Mediated Apoptosis. Bioorg Med Chem 17:1181
15. Ohtsuki T, Hiraka T, Kikuchi H, Koyano T, Kowithayakorn T, Sakai T, Ishibashi M (2009) Flavonoids from *Eupatorium odoratum* with Death Receptor 5 Promoter Enhancing Activity. Heterocycles 77:1379
16. Ohtsuki T, Kikuchi H, Koyano T, Kowithayakorn T, Sakai T, Ishibashi M (2009) Death Receptor 5 Promoter-Enhancing Compounds Isolated from *Catimbium speciosum* and Their Enhancement Effect on TRAIL-Induced Apoptosis. Bioorg Med Chem 17:6748
17. Toume K, Nakazawa T, Ohtsuki T, Arai MA, Koyano T, Kowithayakorn T, Ishibashi M (2011) Cycloartane Triterpenes Isolated from *Combretum quadrangulare* in a Screening Program for Death-Receptor Expression Enhancing Activity. J Nat Prod 74:249
18. Toume K, Nakazawa T, Tahmina H, Ohtsuki T, Arai MA, Koyano T, Kowithayakorn T, Ishibashi M (2012) Cycloartane Triterpenes and Ingol Diterpenes Isolated from *Euphorbia neriifolia* in a Screening Program for Death-Receptor Expression-Enhancing Activity. Planta Med 78:1370

19. Miyagawa T, Ohtsuki T, Koyano T, Kowithayakorn T, Ishibashi M (2009) Cassaine Diterpenoid Dimers Isolated from *Erythrophleum succirubrum* with TRAIL-resistance Overcoming Activity. Tetrahedron Lett 50:4658

20. Miyagawa T, Ohtsuki T, Koyano T, Kowithayakorn T, Ishibashi M (2009) Cardenolide Glycosides of *Thevetia peruviana* and Triterpenoid Saponins of *Sapindus emarginatus* as TRAIL Resistance-overcoming compounds. J Nat Prod 72:1507

21. Ohtsuki T, Miyagawa T, Koyano T, Kowithayakorn T, Ishibashi M (2010) Isolation, Structure Elucidation, and TRAIL-resistance-overcoming Activity of Flavonoid Glycosides from *Solanum verbascifolium*. Phytochem Lett 3:88

22. Ahmed F, Toume K, Sadhu, SK, Ohtsuki T, Arai MA, Ishibashi M (2010) Constituents of *Amoora cucullata* with TRAIL Resistance Overcoming Activity. Org Biomol Chem 8:3696

23. Minakawa T, Toume K, Arai MA, Sadhu SK, Ahmed F, Ishibashi M (2012) Eudesmane-type Sesquiterpenoid and Guaianolides from *Kandelia candel* in a Screening Program for Compounds to Overcome TRAIL-resistance. J Nat Prod 75:1431

24. Clevers H, Nusse R (2012) Wnt/β-Catenin Signaling and Disease. Cell 149:1192

25. Baron R, Kneissel M (2013) Wnt Signaling in Bone Homeostasis and Disease: from Human Mutation to Treatments. Nat Med 19:179

26. Anastas JN, Moon RT (2013) Wnt Signaling Pathways as Therapeutic Targets in Cancer. Nat Rev Cancer 13:11

27. Li X, Ohtsuki T, Koyano T, Kowithayakorn T, Ishibashi M (2009) New Wnt/β-catenin Signaling Inhibitors Isolated from *Eleutherine palmifolia*. Chem Asian J 4:540

28. Mori N, Toume K, Arai MA, Koyano T, Kowithayakorn T, Ishibashi M (2011) 2-Methoxy-1,4-naphthoquinone Isolated from *Impatiens balsamina* in a Screening Program for Activity to Inhibit Wnt Signaling. J Nat Med 65:234

29. Park HY, Toume K, Arai MA, Koyano T, Kowithayakorn T, Ishibashi M (2013) β-Sitosterol and Flavonoids Isolated from *Bauhinia malabarica* Found in a Screening Program for Wnt Signal Inhibitory Activity. J Nat Med 68:242

30. Yamaguchi T, Toume K, Arai MA, Ahmed F, Sadhu SK, Ishibashi M (2012) Phorbol Esters with Wnt Signal-augmenting Effects Isolated from *Excoecaria indica*. Nat Prod Commun 7:475

31. Laursen JB, Nielsen J (2004) Phenazine Natural Products: Biosynthesis, Synthetic Analogues, and Biological Activity. Chem Rev 104:1663

32. Umezawa H, Hayano S, Maeda K, Ogata Y, Okami Y (1950) On a New Antibiotic, Griseolutein, Produced by *Streptomyces*. Jpn Med J 3:111

33. Liu WC, Parker WL, Brandt SS, Atwal KS, Ruby EP (1984) Phenacein - an Angiotensin-Converting Enzyme Inhibitor Produced by a Streptomycete. II. Isolation, Structure Determination and Synthesis. J Antibiot 37:1313

34. Ōmura S, Eda S, Funayama S, Komiyama K, Takahashi Y, Woodruff HB (1989) Studies on a Novel Antitumor Antibiotic, Phenazinomycin: Taxonomy, Fermentation, Isolation, and Physicochemical and Biological Characteristics. J Antibiot 42:1037

35. Kim WG, Ryoo IJ, Yun BS, Shin-Ya K, Seto H, Yoo ID (1997) New Diphenazines with Neuronal Cell Protecting Activity, Phenazostatins A and B, Produced by *Streptomyces* sp. J Antibiot 50:715

36. Abdelfattah MS, Toume K, Ishibashi M (2010) Izumiphenazine A, B and C: Novel Phenazine Derivatives Isolated from *Streptomyces* sp. IFM 11204. J Nat Prod 73:1999

37. Abdelfattah MS, Toume K, Ishibashi M (2011) Izumiphenazine D, New Phenazoquinoline *N*-Oxide from *Streptomyces* sp. IFM 11204. Chem Pharm Bull 59:508

38. Kitamura S, Hashizume K, Iida T, Miyashita E, Shirahata K, Kase H (1986) Studies on Lipoxygenase Inhibitors. II. KF8940 (2-*n*-Heptyl-4-hydroxyquinoline-*N*-oxide), a Potent and Selective Inhibitor of 5-Lipoxygenase, Produced by *Pseudomonas methanica*. J Antibiot 39:1160

39. Kunze B, Hofle G, Reichenbach H (1987) The Aurachins, New Quinoline Antibiotics from Myxobacteria: Production, Physico-Chemical and Biological Properties. J Antibiot 40:258

40. Kato S, Shindo K, Yamagishi Y, Matsuoka M, Kawai H, Mochizuki J (1993) Phenazoviridin, a Novel Free Radical Scavenger from *Streptomyces* sp. Taxonomy, Fermentation, Isolation, Structure Elucidation and Biological Properties. J Antibiot 46:1485

41. Kunigami T, Shin-ya K, Furihata K, Furihata K, Hayakawa Y, Seto H (1998) A Novel Neuronal Cell Protecting Substance, Aestivophoenin C, Produced by *Streptomyces purpeofuscus*. J Antibiot 51:880

42. Pathirana C, Jensen PR, Dwight R, Fenical W (1992) Rare Phenazine L-Quinovose Esters from a Marine Actinomycete. J Org Chem 57:740

43. Abdelfattah MS, Toume K, Ishibashi M (2011) Isolation and Structure Elucidation of Izuminosides A-C: Rare Phenazine Glycosides from *Streptomyces* sp. IFM 11260. J Antibiot 64:271

44. Bock K, Lundt I, Pedersen C (1973) Assignment of Anomer Structure to Carbohydrates through Geminal ^{13}C–^{1}H Coupling Constants. Tetrahedron Lett 14:1037

45. Horinaka M, Yoshida T, Shiraishi T, Nakata S, Wakada M, Nakanishi R, Nishino H, Matsui H, Sakai, T (2005) Luteolin Induces Apoptosis via Death Receptor 5 Upregulation in Human Malignant Tumor Cells. Oncogene 24:7180

46. Sperry J, Bachu P, Brimble MA (2008) Pyranonaphthoquinones – Isolation, Biological Activity, and Synthesis. Nat Prod Rep 25:376

47. Chen Z, Huang H, Wang C, Li Y, Ding J, Sankawa U, Noguchi H, Iitaka Y (1986) Hongconin, a New Naphthalene Derivative from Hong-Cong, the Rhizome of *Eleutherine americana* Merr. et Hyene (Iridaceae). Chem Pharm Bull 34:2743

48. Abdelfattah MS, Toume K, Ishibashi M (2011) New Pyranonaphthoquinones and Phenazine Alkaloid Isolated from *Streptomyces* sp. IFM 11307 with TRAIL Resistance-overcoming Activity. J Antibiot 64:729; corrigendum (2012) J Antibiot 65:277

49. Igarashi M, Chen W, Tsuchida T, Umekita M, Sawa T, Naganawa H, Hamada M, Takeuchi T (1995) 4′-Deacetyl-(−)-Griseusins A and B, New Naphthoquinone Antibiotics from an Actinomycete. J Antibiot 48:1502

50. Bergy ME (1968) Kalafungin, a New Broad Spectrum Antibiotic. Isolation and Characterization. J Antibiot 21:454

51. Ōmura S, Tanaka H, Okada Y, Marumo H (1976) Isolation and Structure of Nanaomycin D, an Enantiomer of the Antibiotic Kalafungin. J Chem Soc Chem Commun:320

52. Mentel M, Ahuja EG, Mavrodi DV, Breinbauer R, Thomashow LS, Blankenfeldt W (2009) Of Two Make One: The Biosynthesis of Phenazines. ChemBioChem 10:2295

53. Gilpin ML, Fulston M, Payne D, Cramp R, Hood I (1995) Isolation and Structure Determination of Two Novel Phenazines from a *Streptomyces* with Inhibitory Activity against Metallo-Enzymes, Including Metallo-beta-Lactamase. J Antibiot 48:1081

54. Abdelfattah MS, Toume K, Ishibashi M (2012) Yoropyrazone, a New Naphthopyridazone Alkaloid Isolated from *Streptomyces* sp. IFM 11307 and Evaluation of its TRAIL Resistance-overcoming Activity. J Antibiot 65:245

55. McNab H (1983) An Analysis of the ^{13}C N.M.R. Spectra of Pyridazin-3-ones and Some Specifically Deuteriated Derivatives. J Chem Soc Perkin Trans 1:1203

56. Bhandari R, Eguchi T, Serine A, Ohashi Y, Kakinuma K, Ito M, Mizoue K (1999) Structure of NG-061, a Novel Potentiator of Nerve Growth Factor (NGF) Isolated from *Penicillium minioluteum* F-4627. J Antibiot 52:231

57. Stefańska B, Arciemiuk M, Bontemps-Gracz MM, Dzieduszycka M, Kupiec A, Martelli S, Borowski E (2003) Synthesis and Biological Evaluation of 2,7-Dihydro-3*H*-dibenzo[*de*,*h*]-cinnoline-3,7-dione Derivatives, a Novel Group of Anticancer Agents Active on a Multidrug Resistant Cell Line. Bioorg Med Chem 11:561

58. Stefańska B, Bontemps-Gracz MM, Antonini I, Martelli S, Arciemiuk M, Piwkowska A, Rogacka D, Borowskia E (2005) 2,7-Dihydro-3*H*-pyridazino[5,4,3-*kl*]-acridin-3-one Derivatives, Novel Type of Cytotoxic Agents Active on Multidrug-Resistant Cell Lines. Synthesis and Biological Evaluation. Bioorg Med Chem 13:1969

59. Arsenault GP (1965) The Structure of Bostrycoidin, a 8-Azaanthraquinone from *Fusarium solani* D$_2$ Purple. Tetrahedron Lett 6:4033

60. Miljkovic A, Mantle PG, Williams DJ, Rassing B (2001) Scorpinone: A New Natural Azaanthraquinone Produced by a *Bispora*-like Tropical Fungus. J Nat Prod 64:1251
61. Moriyasu Y, Miyagawa H, Hamada N, Miyawaki H, Ueno T (2001) 5-Deoxy-7-methylbostrycoidin from Cultured Mycobionts from *Haematomma* sp. Phytochemistry 58:239
62. Koyama J, Morita I, Kobayashi N, Osakai T, Usuki Y, Taniguchi M (2005) Structure–Activity Relations of Azafluorenone and Azaanthraquinone as Antimicrobial Compounds. Bioorg Med Chem Lett 15:1079
63. Barton DHR, Jaszberenyi JC, Liu W, Shinada T (1996) Oxidation of Hydrazones by Hypervalent Organoiodine Reagents: Regeneration of the Carbonyl Group and Facile Syntheses of α-Acetoxy and α-Alkoxy Azo Compounds. Tetrahedron 52:14673
64. Chomcheon P, Sriubolmas N, Wiyakrutta S, Ngamrojanavanich N, Chaichit N, Mahidol C, Ruchirawat S, Kittakoop P (2006) Cyclopentenones, Scaffolds for Organic Syntheses Produced by the Endophytic Fungus *Mitosporic dothideomycete* sp. LRUB20. J Nat Prod 69:1351
65. Lattová E, Perreault H (2009) Method for Investigation of Oligosaccharides Using Phenylhydrazine Derivatization. Methods Mol Biol 534:65
66. Abdelfattah MS, Toume K, Arai MA, Masu H, Ishibashi M (2012) Katorazone, a Novel Yellow Pigment with 2-Azaquinone-Phenylhydrazone Structure Produced by *Streptomyces* sp. IFM 11299. Tetrahedron Lett 53:3346
67. Bauer JD, King RW, Brady SF (2010) Utahmycins A and B, Azaquinones Produced by an Environmental DNA Clone. J Nat Prod 73:976
68. Lyčka A, Jirman J (1995) ^{1}H, ^{13}C and ^{15}N NMR Spectra of Some Anthracenedione Phenylhydrazones. *Dyes Pigments* 28:207
69. Cui HX, Shaaban KA, Schiebel M, Qin S, Laatsch H (2008) New Antibiotic with Typical Plant Anthraquinone Structure Obtained Studying Terrestrial and Marine Streptomycetes. World J Microbiol Biotechnol 24:419
70. Aoki Y, Yoshida Y, Yoshida M, Kawaide H, Abe H, Natsume M (2005) Anthranilic Acid, a Spore Germination Inhibitor of Phytopathogenic *Streptomyces* sp. B-9-1 Causing Root Tumor of Melon. Actinomycetologica 19:48
71. Ma C, Li Y, Niu S, Zhang H, Liu X, Che Y (2011) *N*-Hydroxypyridones, Phenylhydrazones, and a Quinazolinone from *Isaria farinosa*. J Nat Prod 74:32
72. Aida W, Ohtsuki T, Li X, Ishibashi M (2009) Isolation of New Carbamate- or Pyridine-Containing Natural Products, Fuzanins A, B, C, and D from *Kitasatospora* sp. IFM10917. Tetrahedron 65:369
73. Kawabata J, Tahara S, Mizutani J (1978) (2*S*,3*R*)-*trans*-4-*trans*-6-Octadiene-2,3-diol: Structure and Absolute Configuration of a Novel Metabolite of Sorbic Acid. Agric Biol Chem 42:89
74. Hirama M, Shigemoto T, Ito S (1985) Reversal of Diastereofacial Selectivity in the Intramolecular *Michael* Addition of δ-Carbamoyloxy-α,β-unsaturated Esters. Synthesis of *N*-Benzoyl-D,L-daunosamine. Tetrahedron Lett 26:4137
75. Dillon J, Nakanishi K (1975) Absolute Configurational Studies of Vicinal Glycols and Amino Alcohols. I. With Bis(acetylacetonato)nickel. J Am Chem Soc 97:5409
76. Maekawa K, Toume K, Ishibashi M (2010) Isolation of New Fuzanins, Carbamate-Containing Natural Products, from *Kitasatospora* sp. IFM10917. J Antibiot 63:385
77. Shigihara Y, Koizumi Y, Tamamura T, Homma Y, Isshiki K, Dobashi K, Naganawa H, Takeuchi T (1988) 6-Deoxy-8-*O*-methylrabelomycin and 8-*O*-Methylrabelomycin from a *Streptomyces* Species. J Antibiot 41:1260
78. Ishibashi M, Tsukahara K, Toume K (2013) Elmonin. Jap Pat Appl 2013–068062, March 28, 2013
79. Raju R, Gromyko O, Fedorenko V, Luzhetskyy A, Müller R (2013) Oleaceran: A Novel Spiro[isobenzofuran-1,2′-naphtho[1,8-*bc*]furan] Isolated from a Terrestrial *Streptomyces* sp. Org Lett 15:3487

80. Kikuchi H, Ohtsuki T, Koyano T, Kowithayakorn T, Sakai T, Ishibashi M (2010) Activity of Mangosteen Xanthones and Teleocidin A-2 in Death-Receptor Expression Enhancement and Tumor Necrosis-Factor Related Apoptosis-Inducing Ligand Assays. J Nat Prod 73:452

81. Sakai S, Hitotsuyanagi Y, Aimi N, Fujiki H, Suganuma M, Sugimura T, Endo Y, Shudo K (1986) Absolute Configuration of Lyngbyatoxin A (Teleocidin A-1) and Teleocidin A-2. Tetrahedron Lett 27:5219

82. Jin CY, Park C, Cheong JH, Choi BT, Lee TH, Lee JD, Lee WH, Kim GY, Ryu CH, Choi YH (2007) Genistein Sensitizes TRAIL-Resistant Human Gastric Adenocarcinoma AGS Cells through Activation of Caspase-3. Cancer Lett 257:56

83. Ahmed F, Ohtsuki T, Aida W, Ishibashi M (2008) Tyrosine Derivatives Isolated from *Streptomyces* sp. IFM 10937 in a Screening Program for TRAIL-Resistance Overcoming Activity. J Nat Prod 71:1963

84. Hinman JW, Caron EL, Hoeksema H (1957) The Structure of Novobiocin. J Am Chem Soc 79:3789

85. Marfey P (1984) Determination of D-Amino Acids. II. Use of a Bifunctional Reagent, 1,5-Difluoro-2,4-dinitrobenzene. Carlsberg Res Commun 49:591

86. Fujii K, Ikai Y, Mayumi T, Oka H, Suzuki M, Harada K (1997) A Nonempirical Method Using LC/MS for Determination of the Absolute Configuration of Constituent Amino Acids in a Peptide: Elucidation of Limitations of *Marfey*'s Method and of Its Separation Mechanism. Anal Chem 69:3346

87. Tamai Y, Toume K, Arai MA, Hayashida A, Kato H, Shizuri Y, Tsukamoto S, Ishibashi M (2012) Nonactin and Related Compounds Found in a Screening Program for Wnt Signal Inhibitory Activity. Heterocycles 84:1245

88. Beck J, Gerlach H, Prelog V, Voser W (1962) Stoffwechselprodukte von Actinomyceten. 35. Mitteilung. Über die Konstitution der Makrotetrolide Monactin, Dinactin und Trinactin. Helv Chim Acta 45:620

89. Gwak J, Cho M, Gong SJ, Won J, Kim DE, Kim EY, Lee SS, Kim M, Kim TK, Shin JG, Oh S (2006) Protein-Kinase-C-Mediated β-Catenin Phosphorylation Negatively Regulates the Wnt/β-Catenin Pathway. J Cell Sci 119:4702

90. Tamai Y, Toume K, Arai MA, Ishibashi M (2012) Griseoviridin and Cyclic Hydroxamates Found in a Screening Program for Wnt Signal Inhibitor. Heterocycles 86:1517

91. Lin Q, Liu Y, Xi T (2008) The NMR Analysis of Antimicrobial SR-B Compound Produced by Marine Actinomycete N331. Fenxi Ceshi Xuebao 27:911

92. Lee HS, Shin HJ, Jang KH, Kim TS, Oh KB, Shin J (2005) Cyclic Peptides of the Nocardamine Class from a Marine-Derived Bacterium of the Genus *Streptomyces*. J Nat Prod 68:623

93. Takahashi A, Nakamura H, Kameyama T, Kurasawa S, Naganawa H, Okami Y, Takeuchi T, Umezawa H, Iitaka Y (1987) Bisucaberin, a New Siderophore, Sensitizing Tumor Cells to Macrophage-Mediated Cytolysis. II. Physico-Chemical Properties and Structure Determination. J Antibiot 40:1671

94. Seo Y, Cho KW, Rho JR, Mo SJ, Shin J (2001) A New Analog of Antimycin from *Streptomyces* sp. M03033. J Microbiol Biotechnol 11:663

95. Davarani SSH, Fakhari AR, Shaabani A, Ahmar H, Maleki A, Fumani NS (2008) A Facile Electrochemical Method for the Synthesis of Phenazine Derivatives via an ECECC Pathway. Tetrahedron Lett 49:5622

96. Clark TW, Sperry J (2012) Biomimetic Synthesis of Phenazine-1,6-dicarboxylic Acid (PDC). Synlett 23:2827

97. Kometani T, Takeuchi Y, Yoshii E (1983) Pyranonaphthoquinone Antibiotics. 4. Total Synthesis of (+)-Griseusin A, an Enantiomer of the Naturally Occurring Griseusin A. J Org Chem 48:2311

98. Tatsuta K (2008) Total Synthesis and Development of Bioactive Natural Products. Proc Jpn Acad Ser B Phys Biol Sci 84:87

99. Donner CD (2013) Tandem *Michael-Dieckmann/Claisen* Reaction of *ortho*-Toluates—the *Staunton-Weinreb* Annulation. Tetrahedron 69:3747

100. Li Z, Gao Y, Tang Y, Dai M, Wang G, Wang Z, Yang Z (2008) Total Synthesis of Crisamicin A. Org Lett 10:3017
101. Naysmith BJ, Brimble MA (2013) Synthesis of the Griseusin B Framework via a One-Pot Annulation-Methylation-Double Deprotection-Spirocyclization Sequence. Org Lett 15:2006
102. Zhang Y, Wang X, Sunkara M, Ye Q, Ponomereva LV, She QB, Morris AJ, Thorson JS (2013) A Diastereoselective Oxa-*Pictet-Spengle*r-Based Strategy for (+)-Frenolicin B and *epi*-(+)-Frenolicin B Synthesis. Org Lett 15:5566

Genome Mining: Concept and Strategies for Natural Product Discovery

Markus Nett

Contents

1   Introduction

Genomic sciences have undoubtedly transformed the field of life sciences in the past decade. Mapping a genome has become a standard method to address fundamental questions about the physiology and metabolic traits of prokaryotic and eukaryotic organisms. In the area of drug discovery, genomic data are utilized for the identification of potential drug targets (*1*), or for seeking biosynthesis pathways to previously overlooked secondary metabolites (*2*, *3*). While the former approach

M. Nett (✉)

Leibniz Institute for Natural Product Research and Infection Biology,
Beutenbergstr. 11a, Jena 07745, Germany
e-mail: markus.nett@hki-jena.de

has only met limited success to date, the genome-based screening for biosynthesis loci is now considered a routine method in natural product research (*4*). Genome sequencing has unearthed a plethora of untapped chemical diversity in microorganisms and plants (*5*), and has also set the stage for applying rational strategies to recover the predicted molecules (*6*). This development has significantly contributed to the renewed interest of pharmaceutical industry in natural products (*7*, *8*), and has already led to the discovery of an anticancer agent that entered clinical trials (*9*). Another noteworthy example is the antibiotic daptomycin. The correct stereochemistry of this drug, which is approved by the United States Food and Drug Administration (FDA) and by the European Medicines Agency (EMEA), could only be resolved after sequencing of the producing microorganism (*10*).

The term "genome mining", which is frequently encountered in this context, refers to the California gold rush of 1849 and, thereby, to a historical period of emerging opportunities for wealth creation. Likewise, the introduction of highly efficient sequencing platforms and the resulting access to low-cost DNA sequence data can be regarded as a chance for knowledge creation. While early gold mining involved the use of shallow metal pans, in which the gold nuggets were separated from an aqueous soil suspension by gently swirling, the modern day genome mining describes the computational analysis of nucleotide sequence data *via* pattern recognition (*11*). By virtue of this definition, every bioinformatic study aiming at predicting physiological or metabolic properties can be considered as genome mining. In the natural product-related literature, however, the same designation is often used in a more specific sense. Accordingly, genome mining is not limited to the detection of biosynthesis genes by computational methods, but it also involves their functional interrogation, ideally culminating in the elucidation of the associated chemistry.

For the present contribution, it was decided to use the term "genome mining" both for the basic *in silico* analyses, which helps to identify putative biosynthesis genes, and for the emerging chemical or genetic methods that are applied to trace their metabolic products. Following a brief introduction into the different assembly strategies of natural products in Sect. 2, the computational approaches and available software tools for genome mining will be described in Sect. 3. The various experimental strategies and their outcome in terms of discovered molecules will be covered in Sect. 4. Rather than giving an exhaustive overview, selected examples of compounds will be highlighted that belong to the predominant biosynthetic natural product classes. In the final section, the impact of genomics on natural product discovery will be discussed as well as current and emerging trends in this research field.

2   Principles of Microbial Natural Product Biosynthesis

The majority of genome mining studies are conducted in microorganisms, which, of note, does not imply that the metabolomes of plants or other specialist producers, such as sponges or dinoflagellates, are fully explored. An analysis of available

genomic data confirms that there are still a lot of molecules to discover from these drug resources (*12*, *13*). The main reason for the preference of bacteria and fungi in genomic mining probably refers to the clustering of their biosynthesis genes, which is only rarely observed in other organisms (*14*). Open reading frames (ORFs) that are involved in the assembly of a microbial natural product are typically found on a contiguous DNA sequence (*15*). Furthermore, an operon-like gene organization guarantees a coordinate regulation on the transcriptional level (*16*). Due to these features, functionally associated ORFs are more easily assigned in microbial genomes or, in other words, the detection of a single biosynthesis gene is often sufficient to identify an entire metabolic pathway. The *in silico* screening using genetic probes is homology-driven, while functional predictions mostly rely upon biosynthetic precedence. Therefore, it is no surprise that advances in the understanding of natural product biosynthesis are essential for translating genetic or enzymatic information into chemical information. In the following sections, the construction mechanisms and the underlying biochemistry of the four secondary metabolite classes that are primarily targeted in genome mining studies will be introduced, as well as some general guidelines that enable the analysis of unprecedented pathways.

2.1 Polyketides

Polyketides encompass one of the largest groups of microbial secondary metabolites and include many therapeutically significant compounds, such as the antibiotics erythromycin and amphotericin B, the anticancer agent daunorubicin, and the cholesterol-lowering agent lovastatin. All these natural products have in common that they formally originate from the repetitive linkage of acetate and/or propionate units (*17*). Although their building blocks are rather simple molecules, polyketides exhibit a surprising chemical diversity and complexity. The structural manifoldness arises from the concerted action of large biosynthetic enzymes, which are reminiscent of fatty acid synthases (FASs), and which also follow the same assembly strategy. The archetypal type I polyketide synthase (PKS) possesses a modular architecture. Every module represents a discrete functional unit and is typically responsible for carrying out one single elongation step, *i.e.* the attachment of an activated carboxylic acid to the growing polyketide chain. For this purpose, a PKS module must include at least three catalytic domains, namely, an acyltransferase (AT), an acyl carrier protein (ACP), and a β-ketoacylsynthase (KS) domain. The AT domain acts as a biosynthetic gatekeeper. It selects an activated acyl monomer for the chain elongation and tethers the substrate to the PKS protein. The required docking site is located in the ACP domain, which has been post-translationally furnished with a 4′-phosphopantetheine cofactor to provide an active thiol group for the thioesterification of the substrate. The KS domain catalyzes the intrinsic linkage reaction between the monomer extender unit and the acyl precursor by a decarboxylative thio-*Claisen* condensation. To this end, the ACP-bound acyl thioester of the preceding PKS module is initially transthiolated onto a conserved cysteine in the active site of the KS.
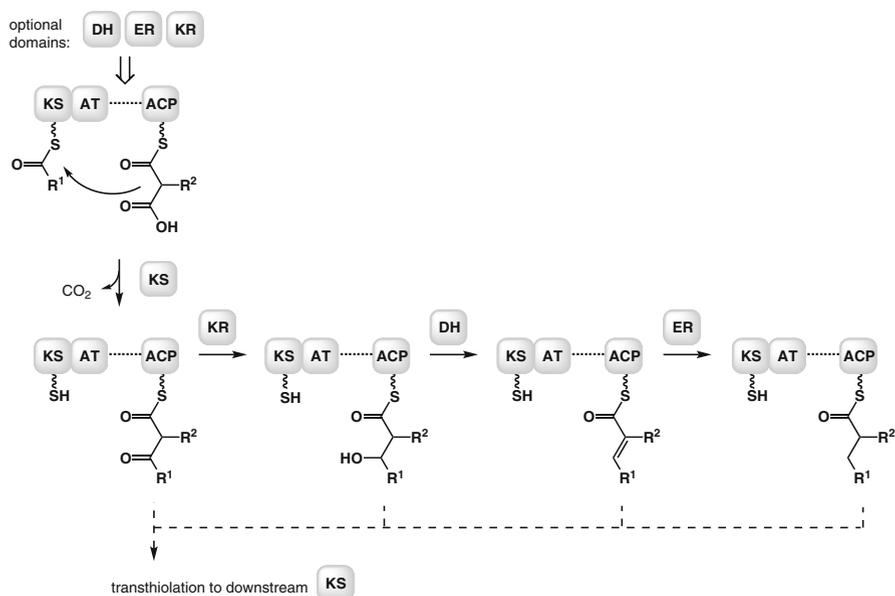
Fig. 1 Acyl chain elongation in polyketide biosynthesis. The roles of individual catalytic domains in a type I PKS are highlighted

The KS-mediated condensation reaction yields a β-ketoacyl thioester intermediate, which can be directly passed onto a downstream PKS module in order to initiate another elongation cycle (Fig. 1, route a). Alternatively, the β-keto functionality can be further processed through the successive action of ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains (Fig. 1, routes b–d).

In the biosynthesis of fatty acids, all intermediates undergo a full reductive cycle, and FASs must therefore feature a complete set of the aforementioned reductive domains. In PKSs, however, the KR, DH and ER domains are optional. Several of these megasynthases are known to lack certain reductive domains or to skip their usage. The varying degree of oxidation significantly enhances the structural diversity of polyketides in comparison to fatty acids and gives rise to many different compound classes, ranging from polyphenols to polyenes, enediynes, macrolides, or polyethers. Once the full chain length is reached in polyketide biosynthesis, the thioester-bound product is detached from the final carrier protein by hydrolysis, lactonization, or reductive release. Subsequently, further modifications of the polyketide core structure can occur, involving tailoring reactions, such as glycosylations, halogenations, and alkylations.

The paradigm described for polyketide assembly was established based upon the biosynthesis of the antibiotic erythromycin (*18*). The erythromycin skeleton is assembled from three multimodular PKSs that are encoded by contiguous genes in the chromosome of the bacterium *Saccharopolyspora erythraea* (Fig. 2). In this example, the number of existing modules exactly matches the number of polyketide building blocks in the final natural product. A sequence-based analysis of the AT
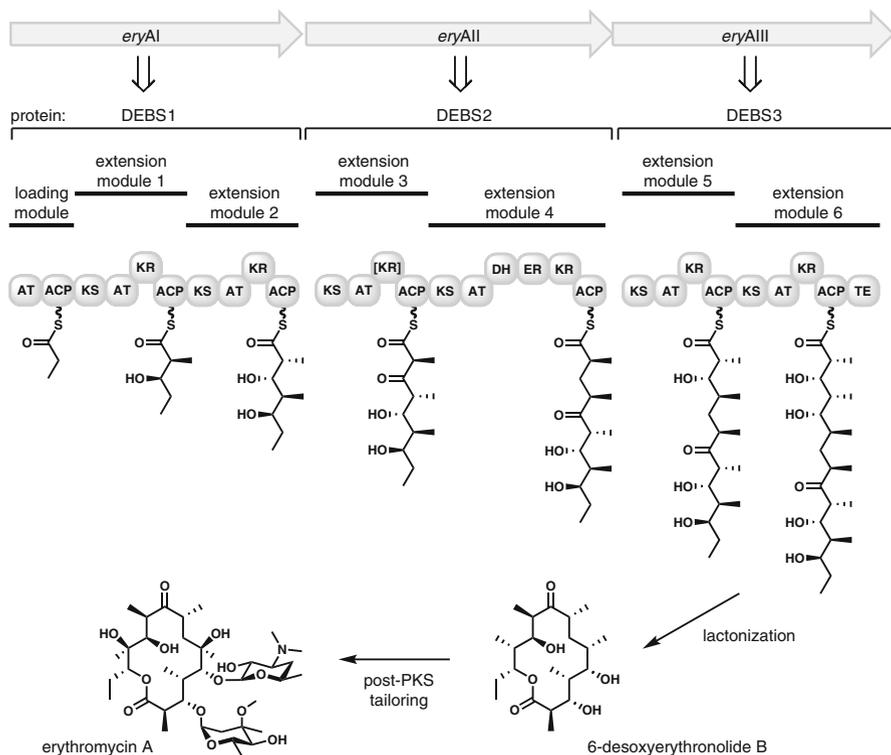
Fig. 2 Genetic and enzymatic basis of erythromycin assembly. The correspondence between domains and biosynthetic transformations illustrates the information transfer in modular PKSs

domains in the erythromycin assembly line also reveals the identity of these mono-mers, among them one propionyl-CoA starter and six methylmalonyl-CoA extender units (*19*). Furthermore, the presence of KR, DH, and ER domains with non-mutated active sites is reflected in the degree of β-keto processing. The close correlation between the PKS machinery, on the one hand, and the structure of the associated natural product, on the other, became known as the principle of colinearity. It is evident that, under the assumption of colinearity, the chemical structure of any polyketide can be predicted from an analysis of its biosynthetic enzymes. Conversely, the constitution of an isolated molecule can also be back translated into the catalytic domains, which constitute the corresponding molecular assembly line.

The general procedure for a polyketide-directed genome mining study would be as follows: It starts with a genome-wide search for the required biosynthesis genes or their deduced proteins. KS and AT domains of characterized PKSs make ideal probes in this context, because their amino acid sequences are usually highly con-served and they also feature a sufficient size, which facilitates their recognition. Once potential candidates for PKSs have been spotted, *e.g. via* homology searches involving the protein-protein Basic Local Alignment Search Tool (BLASTP), the genomic neighborhood of any primary hit is analyzed for associated biosynthesis

genes. Software tools that support the prediction of operons help to assign boundaries of newly recognized biosynthesis gene clusters (*20*). The same is true for large gaps between adjacent ORFs or sudden shifts in their G+C content, which are particularly striking in the genomes of high G+C bacteria (*21*). These changes can be useful markers indicating the transition from a biosynthesis locus to functionally non-related ORFs. Although the identification of the putative cluster boundaries is not required for the prediction of the polyketide core structure, it often provides valuable information on genes that are involved in the preparation of biosynthetic building blocks or in enzymatic tailoring. This knowledge does not only improve the structure proposal, but it can also be used to guide the isolation of the respective natural product. The next step in the genome mining process, after the listing of all cluster genes, is to determine the number of existing PKS modules and their domain architecture. In general, the required data have already been gained during the preceding BLASTP analyses due to the implemented conserved domain search (CDS), yet it is recommended to verify the preliminary results using additional web-based bioinformatic software, such as the Protein family database (Pfam) or the Conserved Domain Architecture Retrieval Tool (CDART). Since the presence of certain domains can have significant effects on the outcome of the structure prediction, care must be taken not to miss small or barely conserved domains. While ER and KR domains are readily identified due to their NADPH-binding sites, the DH and ACP domains of modular PKSs are occasionally overlooked. Subsequently, the correct sequence, in which the modules are used in the assembly process, should be deduced. The two modules, which initiate and terminate the biosynthesis, are comparatively easy to recognize. The former, which is referred to as the loading module, typically lacks a KS domain since no linkage reaction needs to be catalyzed, whereas the latter is often distinguished by a C-terminal thioesterase (TE) or reductive (Red) domain to mediate the offload from the assembly line. In case of the extension modules, the determination of the correct sequence can be more challenging, especially when the transcriptional direction of the corresponding genes switches from the direct to the complementary DNA strand and *vice versa*. If multiple modules are located on a single PKS, their sequence is obvious, starting with the module at the N-terminus and ending with the module at the C-terminus. Prior to translating the enzymatic blueprint into a chemical structure, the biosynthetic building blocks of every elongation step still need to be identified. Unlike FASs, PKSs are capable incorporating monomers that are structurally more elaborate than malonyl-CoA (*22*). Sequence alignments together with site-specific mutations unveiled three distinct regions in every AT domain that govern the respective substrate specificity (*19*). Most useful for bioinformatic predictions is the invariant GHSxGE motif in the active site. If the variable "x" is represented by a branched-chain amino acid, the corresponding AT domain likely has a preference for malonyl-CoA. AT domains that select methylmalonyl-CoA as an extender unit exhibit a glutamine residue in the same position. Once all necessary data have been gathered, the available information is merged into a structure proposal. A scheme of the genome mining process including an illustrative example from the actinomycete bacterium *Frankia* sp. EAN1pec (*23*) is depicted in Fig. 3.

Genome sequence   Example: the actinomycete **Frankia sp. EAN1pec**
(GenBank accession: NC_009921)

① Alignment against conserved PKS domains   BLASTP using a KS query sequence yields 19 significant hits

② Identification of associated genes Assignment of cluster boundaries   4 of the 19 primary hits are encoded by genes that cluster together:

KS   KS   KS   KS

franean1_4838   *4837*   *4836*

③ Analysis of domain architecture   CDS reveals the following domain organization:

Franean1_4838   Franean1_4837   Franean1_4836

KR   DH KR   DH KR   DH KR

Lig ACP KS AT   ACP KS AT   ACP   KS AT   ACP   KS AT ACP

④ Determination of module sequence

extension module 1   extension module 3

loading module   extension module 2   extension module 4

N-terminal ligase (Lig) domain instead of a KS domain

⑤ Substrate specificities of AT domains

AT domain	Active site motifs according to reference [19]			Substrate
module 1	RVEVVQ	GHSQGE	YASH	methyl-malonyl-CoA
module 2	RVEVVQ	GHSQGE	YASH	methyl-malonyl-CoA
module 3	RVEVVQ	GHSQGE	YASH	methyl-malonyl-CoA
module 4	RVDVVQ	GHSQGE	YASH	methyl-malonyl-CoA

⑥ Structure prediction   Linkage of building blocks: Expected reductions due to KR, DH and ER domains:

the starter unit is selected by the ligase domain of Franean1_4838

Fig. 3 The genome mining process, exemplified by the analysis of an orphan PKS locus from the bacterium *Frankia* sp. EAN1pec

Finally, it must be noted, that the colinearity rule is not always applicable. Since the elucidation of erythromycin biosynthesis, several modular or type I PKSs have been discovered that do not obey the canonical rule of polyketide assembly, and which may even exhibit a different enzyme architecture. An example is the occasional absence of AT domains, which are then recruited *in trans* (*24*). Furthermore, two structurally distinct PKS classes can occur in microorganisms. In the so-called type II PKSs, the catalytic and the substrate-binding ACP domains exist as separate proteins that form noncovalently associated complexes to jointly synthesize poly-β-keto chains (*25*). The third group, which is related to the plant-derived chalcone and stilbene synthases, is distinguished by multifunctional, yet modest-sized condensing enzymes that act as homodimers. These type III PKSs do not require ACP domains and typically use free CoA thioesters as substrates (*26*). It is clear that nonmodular and/or iteratively acting PKSs follow different programming rules to those previously discussed. For more detailed informations on this topic the reader is referred to recent reviews (*17*, *24*).

2.2 Nonribosomal Peptides

Akin to polyketides, many peptidic natural products are also assembled by a thiotemplate-based enzymatic logic. Nonribosomal peptide synthetases (NRPSs) are large, modular proteins that possess catalytic domains with functions comparable to those observed in PKSs (*27*). Here, peptidyl carrier protein (PCP) domains inherit the role of substrate binding, while adenylation (A) domains are responsible for substrate recognition. Unlike their enzymatic counterparts in PKSs, however, A domains do not process coenzyme A (CoA)-bound acyl monomers. Instead they activate their substrates in an ATP-driven process and generate the corresponding acyl adenylates. The latter are then transferred onto the phosphopantetheinylated PCP domains. Condensation (C) domains link the thioester-bound precursors to the growing peptide chain *via* amide bond or, more rarely, ester bond formation. Auxiliary domains that are occasionally found in NRPS modules contribute to the structural diversity of nonribosomal peptides. Epimerization (E) domains invert the configuration of the monomers at their α-carbon atoms, while *S*-adenosylmethionine-dependent methyltransferase (MT) domains account for the occurrence of *N*-methylated amino acid residues. Cysteine, serine or threonine residues can be subject to intramolecular cyclization during the linkage reaction, leading to the formation of thiazoline or oxazoline rings, respectively. In these cases, the C domain is replaced by a so-called cyclization (Cy) domain. Further redox adjustments are achieved through the catalytic action of oxidase (Ox) or Red domains. Termination of the elongation process is typically mediated by a thioesterase (TE) domain and can range from simple hydrolysis of the thioester bond to cyclization. A reductive chain release is also possible. As in polyketide biosynthesis, the number of the modules as well as the identity and placement of their domains dictate the composition of the associated natural product. The colinear information transfer of NRPSs is exemplified by the biosynthesis
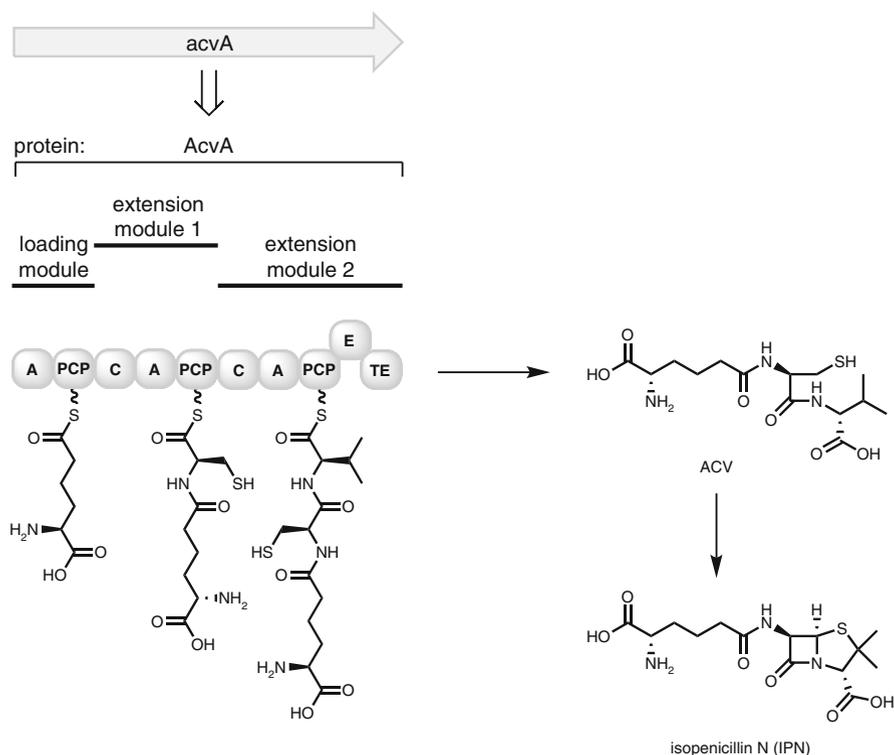
Fig. 4 Genetic and enzymatic of ACV biosynthesis, as an example for a NRPS assembly line

of aminoadipoyl-cysteinyl-valine (ACV), an immediate precursor of β-lactam antibiotics (Fig. 4) (*28*).

For genome mining of NRPS-based pathways, a procedure related to that in Sect. 2.1 (Fig. 3) can be applied. C or A domains, which originate from known enzymes, may serve as primary probes to detect new NRPSs. The prediction of the primary sequence of an NRPS-derived peptide implies the possibility of deducing the substrate specificities of A domains. After the crystal structure of an NRPS from *Bacillus brevis* had been solved (*29*), the key amino acid residues in the binding pocket, which correlate with substrate recognition, were determined *via* sequence alignments and site-specific mutagenesis (*30*, *31*). A signature sequence, which became later known as the 10-amino acid code, was derived from these studies and since then has been used predominantly to analyze the specificity of uncharacterized A domains. It should be noted that the code is highly reliable for bacterial NRPSs, whereas its predictive value seems to be lower for those of fungal origin (*32*). Last, PKSs and NRPSs exhibit a related architecture and enzymology, and it is therefore not surprising that Nature mixes and matches both protein classes. Although the resulting hybrid assembly lines switch from PKS to NRPS interfaces and *vice versa*, they allow for *in silico* predictions of their metabolic products due to a compatible sequential programming (*33*).

2.3 Ribosomally Synthesized and Post-translationally Modified Peptides

Although ribosomally synthesized and post-translationally modified peptides are ubiquitous in Nature and can be found in all three domains of life, their significance and number were both severely underestimated for many years. A standardized nomenclature for these natural products has only recently been introduced and, accordingly, they are now briefly referred to as RiPPs (*34*). RiPPs derive from precursor peptides that typically consist of 20–110 amino acid residues. Upon more or less extensive post-translational modifications of a core region, an N-terminal leader sequence is removed proteolytically to yield the mature peptide, which is then usually transported out of the producing cell (Fig. 5). Some rare cases are known, where precursor peptides lack a leader sequence and rather exhibit a C-terminal "follower" peptide sequence that is cleaved in the biosynthesis. Further exceptions from the archetypal organization include the presence of a C-terminal recognition sequence or the fusion of the leader peptide to a signal sequence, as observed in precursor peptides of eukaryotic RiPPs (*34*).

The awareness that several peptidic natural products are actually of ribosomal and not of NRPS origin is due to the analysis of genomic data. The biosynthesis
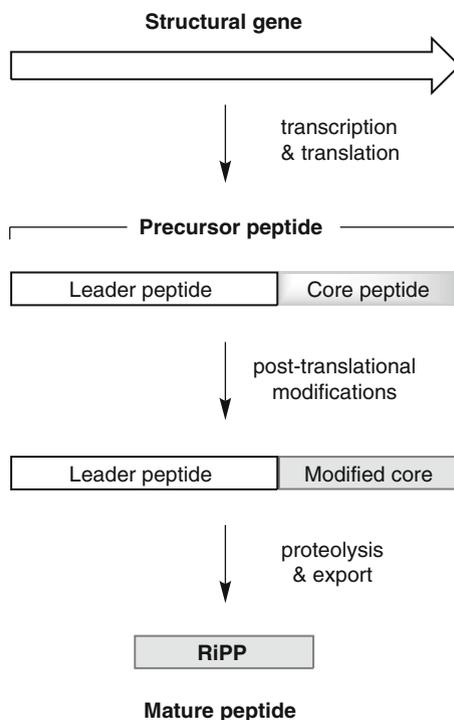
Fig. 5 General pathway for RiPP biosynthesis

genes of the corresponding compounds can be readily identified by searching trans-
lated genome sequences for the possible core peptides (*35*). Conversely, genomic
mining approaches for the discovery of novel RiPPs turn out to be more challeng-
ing. Short open reading frames that code for potential precursor peptides are diffi-
cult to assign, unless their genetic context allows further conclusions. It is thus not
surprising that genome mining studies typically target the conserved enzymes
accounting for the post-translational modifications and for which the open reading
frames are located closely to the structural genes. Illustrative examples in this
respect are the enzymes that generate the distinctive structural motifs of RiPPs
belonging to the lanthipeptide family.

Lanthipeptides feature thioether moieties that crosslink the β-carbon atoms of
either two alanine residues or one alanine and one α-aminobutyrate residue. The
formation of these thioether linkages requires two successive post-translational
modifications. Initially, select serine and/or threonine residues of the core peptide
are dehydrated to their corresponding dehydro amino acids, before intramolecular
Michael-type additions of cysteine thiol groups accomplish the ring closures
(Fig. 6). Contingent upon the involved enzymes, four different classes of lanthipep-
tides can be distinguished (*36*). For class I lanthipeptides, the dehydration and cycli-
zation reactions are catalyzed by discrete proteins, whereas the biosyntheses of
classes II, III, and IV lanthipeptides resort to multifunctional enzymes that are only
partially conserved. By exploiting their unique motifs, it becomes possible to screen
genomic sequences for gene clusters that are associated with the biosynthesis of
individual lanthipeptide classes (*37*).
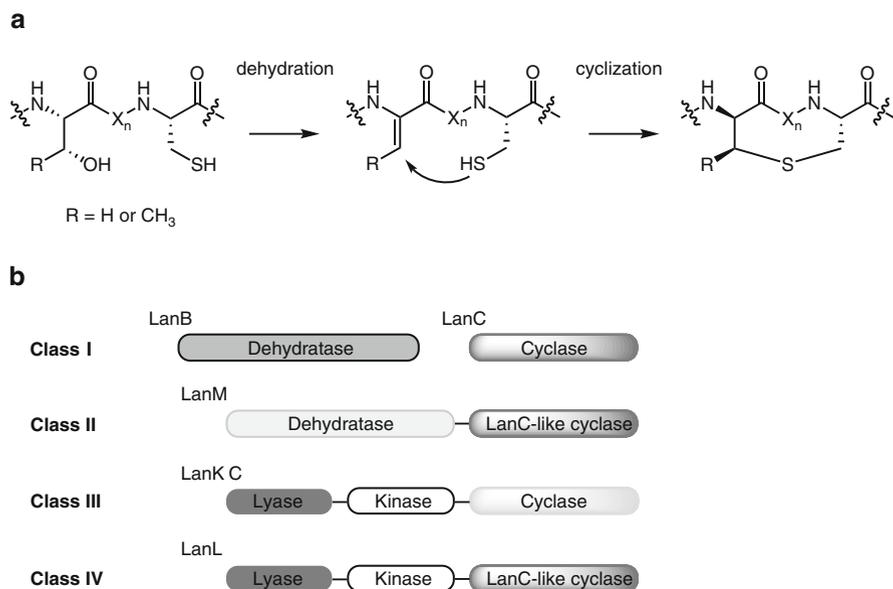
Fig. 6 Post-translational modification of precursor peptides by lanthipeptide synthetases (**a**), and
illustration of the catalytic domains found in different classes of this enzyme family (**b**)

2.4 Terpenes

Terpenes are considered to constitute the largest family of biosynthetically related, naturally occurring compounds. Although they are particularly noteworthy in plants and fungi, reports on bacterial terpenes have accumulated in recent years (*38*). All terpenes originate from the two building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). The two isoprene units can be generated by distinct pathways. While eukaryotes and archaea synthesize IPP predominantly *via* the mevalonate (MVA) pathway, bacteria are often restricted to the methylerythritol phosphate (MEP) pathway, which is also active in the chloroplasts of higher plants and in green algae. Through sequential 1′-4 condensation of IPP to a DMAPP starter molecule a linear polyprenyl chain is assembled. Subsequently, most terpenes undergo a series of cyclizations, which can be accompanied by structural rearrangements and even elimination of carbon atoms. The resulting compounds are often further elaborated due to hydroxylations, methylations, and/or glycosylations (Fig. 7).

Even though terpene biosynthesis enzymes generally display only low conservation in terms of primary amino acid sequence similarity, prenyltransferases and terpene cyclases have been demonstrated to be highly useful for genome mining studies. All members of the cyclase family and also the isoprene linking prenyltransferases share two highly conserved motifs, which are required for cooperative binding of three Mg^{2+} ions together with the pyrophosphate moieties of their substrates (*39*, *40*). According to crystallographic studies the two motifs are located at
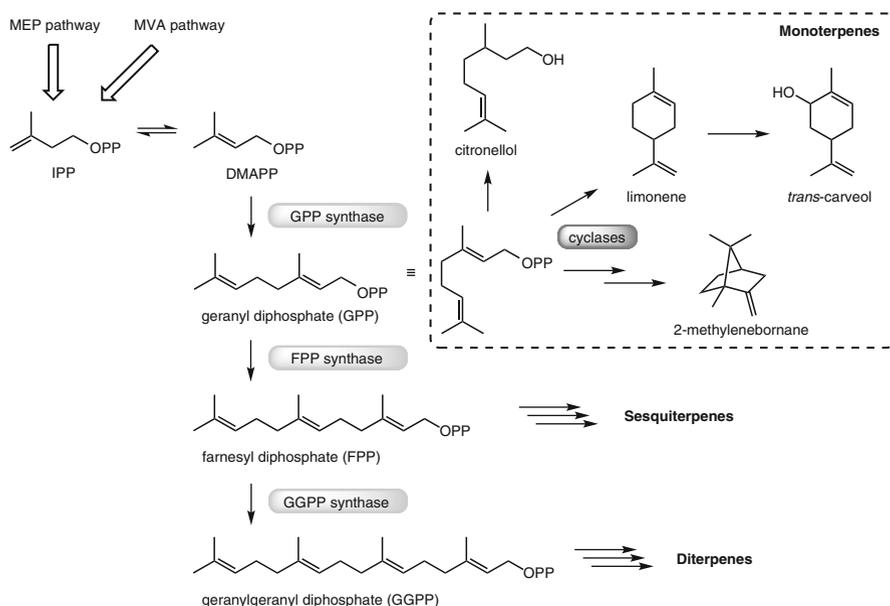
Fig. 7 Principles of isoprenoid biosynthesis

opposite sides of the rim of the respective deep active site cavity. They feature an acid-rich region, DDxxD (or DxDD in case of diterpene cyclases), and a downstream NSE/DTE triad. Profile Hidden *Markov* Models have been instrumental in identifying proteins that contain these signatures sequences and could thus be used to unveil several hidden terpene pathways (*41*). On the other hand, bioinformatic analyses are still limited in assigning specific functions to newly recognized terpene cyclases. Unlike the pathways leading to polyketides, nonribosmal peptides, or RiPPs, predictions of terpene structures are hardly possible, and attempts to trace the products of orphan terpene cyclases are most often limited to the genetic approaches that will be discussed in Sect. 4.2.

3   Bioinformatic Software Tools

The genome mining process can be roughly divided into two successive steps (*cf.* Sect. 2.1). First, target proteins are identified from translated genomic sequences based upon their homology to characterized representatives or due to the presence of conserved motifs. In the second step, the detected hits are subject to detailed analyses, which may also involve their genomic vicinity, in order to assign a possible function. Basic Local Alignment Search Tool (BLAST) and the profile Hidden *Markov* Model (HMM) suite of tools are most commonly employed for conducting the initial screening. While BLAST searches will retrieve any sequence having at least a short region in common with the query sequence, HMMs are statistical models that are applied for pattern recognition. In the context of genome mining, HMM analyses are particularly useful for identifying proteins that contain a specific domain. Although BLAST and HMMER as well as the subsequent bioinformatic studies can be carried out manually, it is evident that computational approaches offer numerous opportunities for automation. In the past decade, several software programs and web-based platforms have been developed to facilitate the prospecting of microbial genomes for biosynthesis gene clusters and also to predict the structures of the associated secondary metabolites (*42*). A survey of the bioinformatic tools, which are currently available, is given in Table 1.

SEARCHPKS and its successor NRPS-PKS were the first computational online tools to predict domain organization and substrate specificities of PKSs and NRPSs, respectively (*43*, *44*). Both software programs pursue a knowledge-based strategy for protein analysis. The substrates of A domains are assigned using the 10-amino acid code (*30*), whereas the specificity of AT domains is traced to 13 active site residues. These sites had been identified in preceding alignments and modeling experiments (*45*). Recently, the NRPS-PKS data training set was significantly extended, improving the accuracy of AT substrate predictions. Moreover, the program was incorporated into the SBSPKS software package, which also permits 3D modeling of PKS catalytic domains and evaluation of domain-specific protein-protein interactions between PKSs (*46*). The latter can be used to determine the order of substrate channeling in an assembly line that is composed of multiple PKSs, optimizing the structural prediction of the resulting polyketide.

Table 1 Software tools for the analysis of biosynthesis enzymes and automated genome mining

Software	Target proteins	Domain specifity	Recognition of inactive domains	Structure prediction	Gene cluster prediction
NRPS-PKS	PKS, NRPS	AT, A			
ASMPKS	PKS	AT		(+)	+/−
ClustScan	PKS	AT, KR	+	+	
NP.searcher	PKS, NRPS	AT, A		+	+/−
NRPSpedictor2	NRPS	A			
NRPSsp	NRPS	A			
NRPS/PKS substrate predictor	PKS, NRPS	AT, A			
NaPDoS	PKS, NRPS				
CLUSEAN	PKS, NRPS	AT, C, A	+/−	+	
SMURF	PKS, NRPS, PT				+
PKMiner	PKS(II)			+	
SEARCHGTr	GT	GT			
BAGEL2	RiPP precursor peptide			(+)	(+)
antiSMASH 2.0	all biosynthetic enzymes	AT, KR, C, A	+/−	+	+

The software platform ASMPKS is confined to the analysis of modular type I PKSs (*47*). It provides four discrete features: The user can (i) search microbial genome sequences for known and unknown polyketide gene clusters, or (ii) analyze the protein sequence of a PKS for the presence of catalytic domains. In the latter case, ASMPKS also predicts the substrate preference of AT domains. Furthermore, the software integrates (iii) a tool to visualize the planar carbon skeleton of a polyketide from a manually inserted modular assembly line and (iv) a database that includes both the chemical structures of selected polyketides and the domain architecture of their biosynthesis enzymes.

ClustScan is a client–server based application that allows the direct processing of nucleotide sequence data. Following an automatic translation in all six possible reading frames, the resulting protein sequences are screened for conserved PKS domains by HMMER analyses (*48*). Using a fingerprint method, the software deduces the substrate being selected by an AT domain. Furthermore, it considers the KR-derived stereochemistry in the prediction of polyketide structures. While NRPS-PKS and ASMPKS cannot distinguish between active domains and those that have lost their functionality due to active site mutations, *ClustScan* specifically searches for such subtle changes and integrates the results in the structure proposal. Another useful feature is the possibility to manually adjust and customize the automatically generated predictions.

The open source software NP.searcher offers a unique feature when compared to other bioinformatic tools that are available for analyzing PKSs and NRPSs. On the basis of biosynthetic and chemical reasoning, it calculates the different structures that may possibly arise from one single assembly line (*49*). Similar to ASMPKS,

the identification of biosynthesis genes as well as the subsequent domain assignment rely on homology-driven BLAST analyses. BLAST is also used to extract the specificity-conferring signature residues of AT and A domains. From the sequence of predicted substrates, a linear natural product is deduced. This hypothetical intermediate structure is then systematically altered. The possible modifications are limited, however, to intramolecular reactions of functional groups, or the modifications must be related to accessory proteins that are encoded in the cluster. In this way, NP.searcher can provide multiple structural proposals for every PKS- and/or NRPS-derived natural product.

As stated before, the substrate specificity of select biosynthesis domains can be inferred from the manual or computational identification of signature sequences (*30*, *31*). An alternative approach to accomplish the same task involves the usage of machine learning methods, as exemplified by the web server NRPSpredictor (*50*). The training data for the required Support Vector Machine (SVM) included a total of 397 A domains with known specificities. From each A domain, a selection of 34 amino acids lining the active site was extracted. These residues were correlated with individual properties that likely contribute to substrate binding, such as volume or the existing number of hydrogen bond donors. The resulting physico-chemical fingerprint was proven to enable a highly accurate substrate prediction in case of bacterial NRPSs. Recently, a new version of the popular prediction software was released. NRPSpredictor2 is based on a larger training set and also provides a predictor that was specifically designed for the analysis of fungal A domains (*51*).

Additional tools for determining the substrates of modular enzymes have recently been described (*52*, *53*). Both NRPSsp and the NRPS/PKS substrate predictor use classification methods that exploit substrate-specific HMMs and do not require the recognition of signature sequences. A similar procedure had previously been reported in the literature, but was not implemented into publicly available software (*54*). Here, substrate specific residues were defined using a quantiative evolutionary trace analysis. In this way, every residue within a protein sequence is ranked according to its relative importance. The latter is calculated from the number of partitions in a phylogenetic tree that are necessary to observe conservation at a specific position. Such traced sites were used for training HMM profiles, which then provided highly accurate predictive markers.

NaPDoS is a bioinformatic tool that exclusively utilizes phylogenetic reconstruction for the analysis of PKS and NRPS biosynthesis genes (*55*). Uploaded sequences are initially scanned for the presence of putative KS or C domains. Upon the detection of potential candidates, the corresponding sequences are extracted, trimmed and aligned with those of characterized KS and C domains. The resulting alignments are then used to compute phylogenetic trees. Clustering patterns, which are derived from KS domain phylogenies, allow the user to distinguish between iterative and non-iterative PKSs. Furthermore, they can discriminate between AT-containing and AT-lacking PKS modules, of which the latter are known as *trans*-AT systems (*24*). NaPDoS is a useful platform to assess the expectable biosynthetic diversity encoded in a draft genome or in a metagenome, although it does not provide structure-related information.

CLUSEAN was developed for the in-depth analysis of single biosynthesis gene clusters (*56*). It features standard sequence annotation as well as functional assignment of PKS and NRPS domains, integrating among others the A domain specificity prediction of NRPSpredictor (*50*). CLUSEAN analyses yield information on the domain architecture, on conserved sequence motifs, and on the biosynthetic bricks of polyketide and nonribosomal peptide natural products. For the mapping of entire secondary metabolite clusters in fungal genomes, a specific web-based software tool named SMURF has been launched (*57*). The identification of core biosynthesis as well as accessory genes relies on HMM searches for conserved protein domains in the translated genome sequence.

Recently, a function-oriented, computational sequence analysis involving HMM calculation and machine learning methods has led to the identification of specific domain classifiers that can be linked with chemotype prediction rules for the iteratively acting type II PKSs. These rules have been integrated into the software PKMiner (*58*), which actually represents the only genome mining tool for the structural prediction of type II PKS-derived aromatic polyketides.

Comparatively few bioinformatic software packages have been developed for the analysis of biosynthesis enzymes other than PKSs and NRPSs. SEARCHGTr is one such example, revealing the substrates of glycosyltransferases (GT) and proposing possible acceptor molecules (*59*). The interactive web server is thus suited for analyzing post-assembly line modifications of polyketide and nonribosomal peptide natural products.

The genes encoding the structural backbones of RiPPs are typically small and poorly conserved. Sometimes they are even overlooked in the genome annotation process. The web-based tool BAGEL uses a strategy for unveiling RiPP genes that is largely based on their genomic context (*60*). Structural genes are often accompanied by accessory ORFs, which are required for post-translational modifications, transport and immunity. In general, these genes are larger than the RiPP genes, and they are also much better conserved, which makes them more easy to identify by sequence alignments. Clustering of such accessory genes is determined *via* distance profiling in BAGEL and serves as an important factor to estimate the likeliness of a nearby RiPP gene. Possible candidates for RiPP genes, which have been identified previously through searches in a bacteriocin database and a motif database, respectively, are ranked according to the presence (or absence) of putative accessory genes. Additional weighting factors include the size of the structural gene, specific motifs in the translated sequence and physical properties. The prediction accuracy of this complementary approach was further improved in BAGEL2, which resorts to a larger knowledge database and implements an advanced classification algorithm (*61*).

Last, antiSMASH represents a comprehensive bioinformatic tool for automated genome mining. This software detects the large number of pathways leading to secondary metabolites (*62*). In case of polyketide and nonribosomal peptide gene clusters, antiSMASH even predicts the chemical structures of their products. For this purpose, the software combines established methods for the prediction of substrate specificities (*45*, *50*, *51*, *54*) and stereochemistry (*48*). The module sequence of an assembly line is estimated from PKS docking domain sequence residue matching (*46*), and in cases where this is not possible, colinearity is assumed. A valuable

feature of antiSMASH is the annotation of entire gene clusters. In addition to core biosynthesis genes, such as those coding for PKSs or terpene cyclases, the software also searches for accessory ORFs by cross-checking the genomic vicinity of the primary hits against a database. The latter includes complete sets of genes from known biosynthesis clusters. This comparative approach enables the identification syntenic loci that may yield the same chemistry. The freely available antiSMASH is typically accessed on a web server, but can also be downloaded as a stand-alone version with a Java graphical user interface. Recently, the software has been upgraded. The new version of antiSMASH provides structural proposals for lanthipeptides (see Sect. 2.3) and identifies subclusters that are involved in the preparation of specific biosynthetic building blocks (*63*).

4 Discovery of Natural Products from Orphan Pathways

The exploration of microbial genomes led to the discovery of numerous biosynthetic gene clusters that could not be associated with the production of known compounds. It has been proposed to refer to such loci as "orphan gene clusters" by analogy to orphan receptors, for which the ligands and functions still need to be identified (*6*). In recent years, several methods were described in order to disclose the compounds that hide behind orphan clusters. In the following paragraphs, these different methods will be presented.

4.1 Structure- and Bioactivity-Guided Strategies

Screening approaches that target chemical and/or biological properties of a natural product have in common their broad applicability, since the organism to be investigated must not be genetically tractable. On the other hand, the success of such techniques strongly depends on the reliability and accuracy of the preceding bioinformatic predictions, be it structural features or biological activities.

4.1.1 UV-Based Methods

The screening of microbial culture extracts for metabolites with characteristic UV profiles is a straightforward approach for the discovery of natural products associated with orphan pathways. In the case of polyketides, the bioinformatic analysis of their biosynthesis enzymes might indicate the presence of an olefinic partial structure. Polyenes are known to absorb UV light at specific wavelengths, which can be even predicted from their structures by applying empirical rules, such as those defined by *Woodward* (*64*). Likewise, bioinformatic studies could reveal the incorporation of a shikimate-derived building block in a natural product. Aromatic amino acids and many of their derivatives are common substrates of NRPSs, and they also

possess distinctive UV absorbance maxima. Although UV-based strategies can only be used for retrieving compounds with a chromophore, a significant number of natural products have been discovered in this way (Fig. 8).
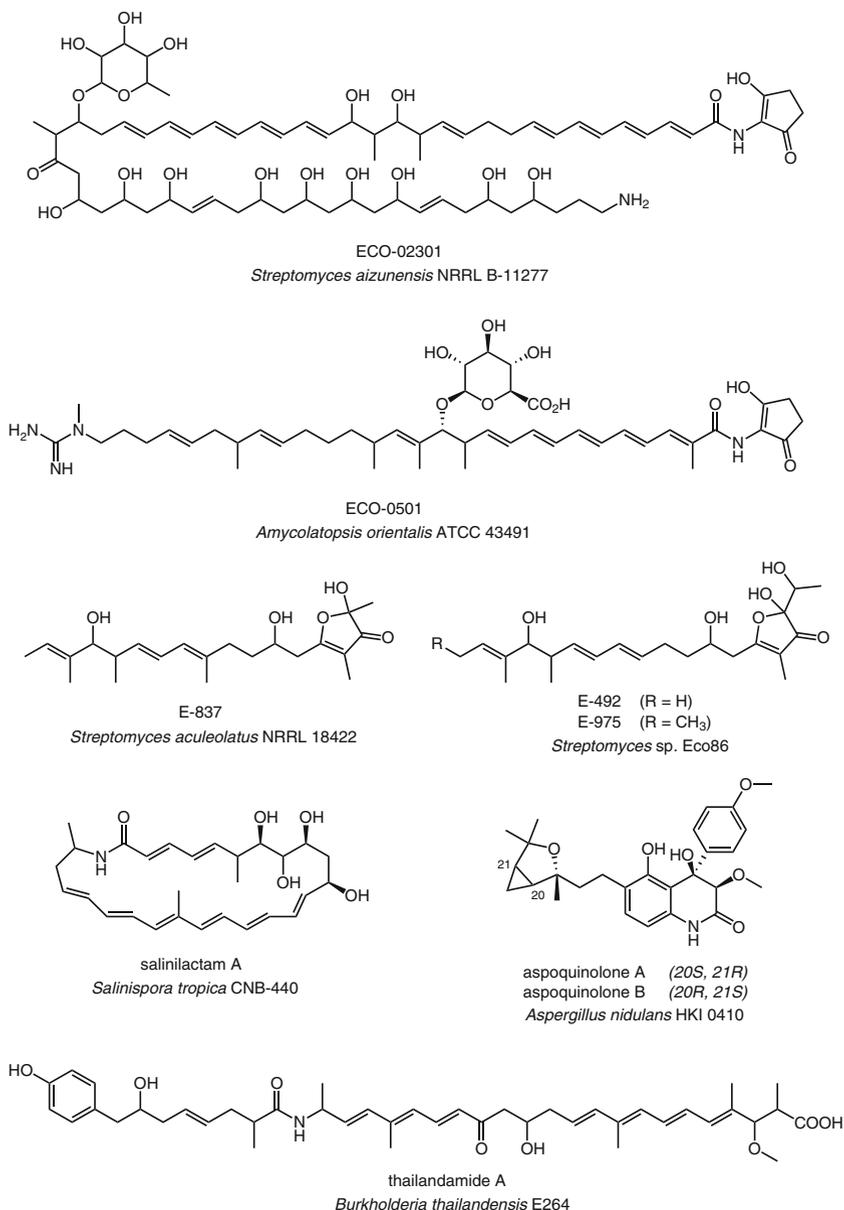
Fig. 8 Examples of natural products identified by their predicted UV profiles. The organisms from which the molecules were isolated are given in parentheses

Ecopia BioSciences spearheaded this inexpensive technique and reported the isolation of several polyene antibiotics from actinomycete bacteria, including the antifungal agent ECO-02301, the anti-MRSA drug ECO-0501 as well as the electron transport inhibitors E-837, E-492, and E-975 (*65–67*). In all of these studies, the genomics-guided screening process was further backed by mass spectrometric analyses, and structure elucidation of the isolated compounds was mainly based on the bioinformatic predictions. Another impressive example was the finding of a new class of polyene macrolactams in the marine actinomycete *Salinispora tropica* CNB-440 (*68*). This bacterium was already known as a source of bioactive natural products (*69*), but the identification of a large modular PKS-based assembly line in its genome and the subsequent structure prediction were key to the discovery of the previously overlooked salinilactams. It is interesting to note that the NMR-based structure elucidation of the salinilactams also helped to resolve and properly assemble the highly repetitive DNA sequences of the PKS cluster, thereby contributing to the closure of the *S. tropica* genome sequence. The thailandamides are polyenes, which are produced by the bacterium *Burkholderia thailandensis* E264, albeit only during its early growth phase. The discovery of these natural products was extremely challenging as a result of their chemical instability, and it became only possible due to an accurate prediction of their chromophore, which consists of several conjugated double bonds (*70*). UV-based methods were also successfully applied for genomics-inspired drug discovery in fungi. Annotation of the *Aspergillus nidulans* genome revealed the presence of several anthranilate synthase genes, provoking efforts to identify aromatic, nitrogen-containing metabolites in extracts of the ascomycete. This approach led to the isolation of aspoquinolones A-D, a class of aromatic polyketides with unprecedented terpenoid side chains (*71*).

4.1.2 NMR-Based Methods

Despite its highly discriminatory power, NMR spectroscopy has only rarely been used to screen microbial extracts for the products of orphan pathways. Possible reasons for the lack of corresponding studies are the low sensitivity of this spectroscopic technique and the challenges associated with the interpretation of spectra that contain the signals of multiple components. Both issues can be circumvented by labeling the target molecule through the biosynthetic incorporation of ^{13}C or ^{15}N enriched precursors. Subsequently, signals of the labeled compound can be easily distinguished in NMR spectra due to the naturally low abundance of ^{13}C and ^{15}N isotopes. The discovery of the NRPS-derived orfamides (Fig. 9) was a proof-of-concept for this method, which became known as the genomisotopic approach (*72*). In the context of natural product discovery, the feeding with isotopically labeled compounds in combination with NMR analyses seems to be generally applicable in microorganisms, presuming a correct prediction of the biosynthetic precursors. Problems may arise, however, when the labeled precursor is heavily metabolized and diverted into different pathways. To avoid an unspecific labeling of the entire metabolome, the usage of uncommon precursors (*e.g.* nonproteinogenic amino

orfamide A
Pseudomonas fluorescens Pf-5

taiwachelin
Cupriavidus taiwanensis LMG19424
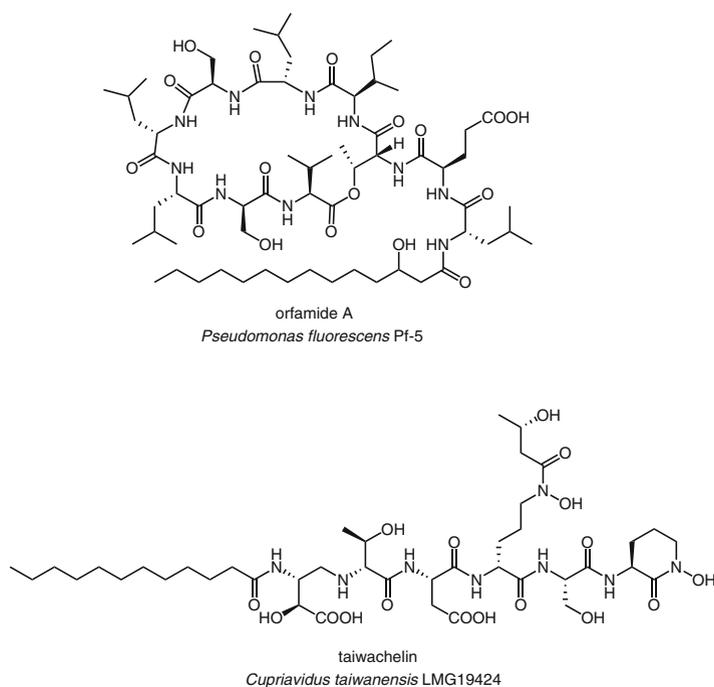
Fig. 9 Natural products that were discovered by NMR-based genome mining methods

acids) is hence recommended. The second example for genomics-inspired natural product discovery by NMR was recently reported. The genome of the bacterium *Cupriavidus taiwanensis* LMG19424 was found to harbor a gene cluster encoding a hydroxamate-containing natural product. Hydroxamates are known to possess highly distinctive ^{15}N NMR resonances, and the authors were able to identify the respective signals in crude extracts of *C. taiwanensis via* ^{1}H,^{15}N HMBC measurements. NMR-guided fractionation then yielded the new lipopeptide taiwachelin (*73*). Other cases, in which a NMR-based screening might be applicable without the need for previous isotopic labeling, could include the search for the rare secondary metabolites containing fluorine or phosphorus atoms, such as phosphonate and phosphinate antibiotics.

4.1.3 MS-Based Methods

Although the genomisotopic approach was designed originally to facilitate the NMR-guided discovery and isolation of natural products, it is obvious that the same methodology can also easily be combined with MS-based analyses. The main advantage of MS when compared to NMR spectroscopy is the increased sensitivity, which also enables inverse labeling experiments. Here, the microorganism of interest is grown in
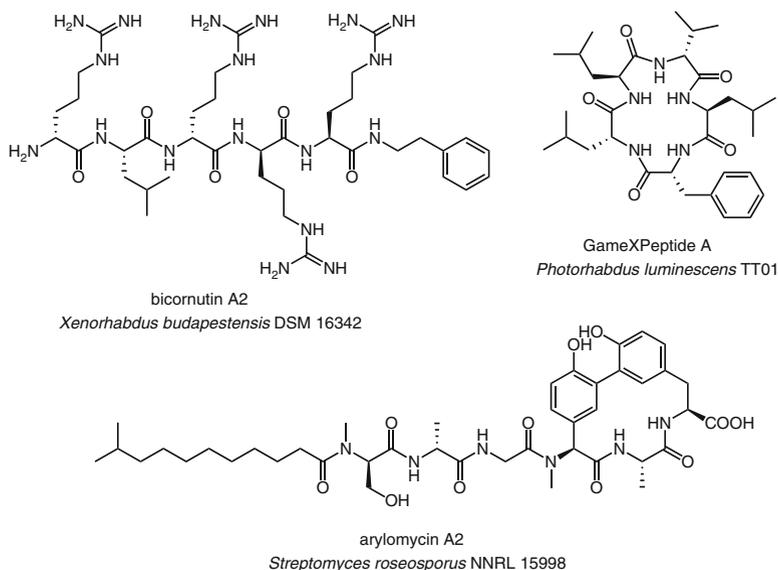
Fig. 10 Natural products that were discovered by MS-based genome mining methods

a fully labeled ^{13}C or ^{15}N medium, which has been supplemented with a non-labeled precursor. The precursor incorporation into a specific molecule is then accompanied by a shift to lower masses. In *Photorhabdus* and *Xenorhabdus* bacteria, the isotopic labeling approach promoted the MS-based structure elucidation and configurational assignment of novel cyclopeptides named GameXPeptides (Fig. 10) (*74*).

A non-isotope MS method emerged from proteomics. In proteomics, the identification of a protein typically involves sequencing by means of tandem MS (MS^n) analyses. The MS^n spectrum reveals diagnostic fragment ions that differ in mass by the amino acid monomers of the corresponding peptide sequence. Based upon the masses of the 20 proteinogenic amino acids, a series of mass shifts can be translated into an amino acid sequence. Except for a few cases, where two distinct amino acids exhibit the same mass (*e.g.* leucine and isoleucine), such predictions are highly specific. Even though most MS^n analyses yield truncated protein sequences, these partial sequences are still sufficient to allow an unequivocal identification by querying them against a protein database. This established methodology has recently been adapted to enable the discovery of secondary metabolites. In the peptido-genomics approach, a series of MS^n-derived mass shifts is traced back to the genes that may account for the biosynthesis of a peptidic natural product with a matching amino acid sequence (*75*). As previously discussed, the amino acid backbone of a RiPP is defined by its precursor peptide, while the core structure of a nonribosomal peptide can be deduced from the amino acid specificity that is encoded on the associated modular synthetases (see Sect. 2.2 and 2.3). Due to post-translational modifications or the incorporation of nonproteinogenic amino acids, the possible mass shifts in natural products exceed those observed in proteins. These biosynthesis peculiarities

must be considered when the amino acid sequence of a natural product is predicted from a MS^n spectrum. In the case of RiPPs, an amino acid sequence featuring at least five successive residues should be available in order to facilitate the linkage of the observed chemotype with its corresponding genotype. The query space of RiPPs includes the complete six-frame translation of the genome, and shorter search tags may simply yield too many candidate precursor peptides. Conversely, the query space is restricted to the comparatively small number of NRPS gene clusters when searching for nonribosomal peptides. Here, short sequence tags of just two amino acids can be sufficient for a successful assignment, as exemplified by the detection of the arylomycins in *Streptomyces roseosporus* ([76]). While the arylomycins represented already described natural products, a total of nine, previously unknown RiPPs were identified by applying the peptidogenomics approach in a single study ([75]). Although the proposed structures of these compounds were not verified by NMR or X-ray analyses, this result underscores the high-throughput capabilities of MS^n-guided genome mining, which may be even expandable to glycosidic natural products. Recently, a complementary strategy to determine the structures of peptide natural products with limited backbone fragmentation has been described. This method relies on the recognition of neutral loss fragmentation pattern, and was used for the identification of several arginine-rich peptides, such as bicornutin A2 ([77]).

4.1.4 Assay-Based Methods

The *in silico* analysis of biosynthesis gene clusters often yields valuable information on structural features of their associated natural products. In some cases, this knowledge can be exploited in order to screen the culture broth of a microorganism for the occurrence of a specific metabolite (Fig. 11). A well-known example involves the application of the chrome azurol S (CAS) assay to detect the release of siderophores ([78]). Siderophores are Fe(III)-scavenging molecules that are used by microorganisms to maintain their cellular iron homeostasis. Their production responds to the iron concentration of the environment and can thus be intentionally triggered in the laboratory. The subsequent CAS activity-guided isolation has proven to be a reliable method to retrieve the products of siderophore pathways, as exemplified by the fuscachelins ([79]).

From a pharmaceutical perspective, it would be desirable if genome mining allowed the investigator to correlate the structures of predicted compounds with their biological activities or molecular targets. This would permit fermentation studies to be focused on chemical entities that possess preselected pharmacological properties. To date, however, such activity-guided genome mining has only been successful when the hypothetical structure contained a known pharmacophore. An example in this context is the search for enediyne pathways in actinomycetes. Enediynes are potent anticancer agents that induce irreversible DNA damages *via* a radical mechanism. The activity of enediynes is due to the presence of a highly reactive warhead chromophore, for which the biosynthesis involves a conserved set of enzymes including an iterative type I polyketide synthase ([80]). In the case of

fuscachelin A
Thermobifida fusca ATCC 2773

lomaiviticin C
Salinispora tropica CNB-440
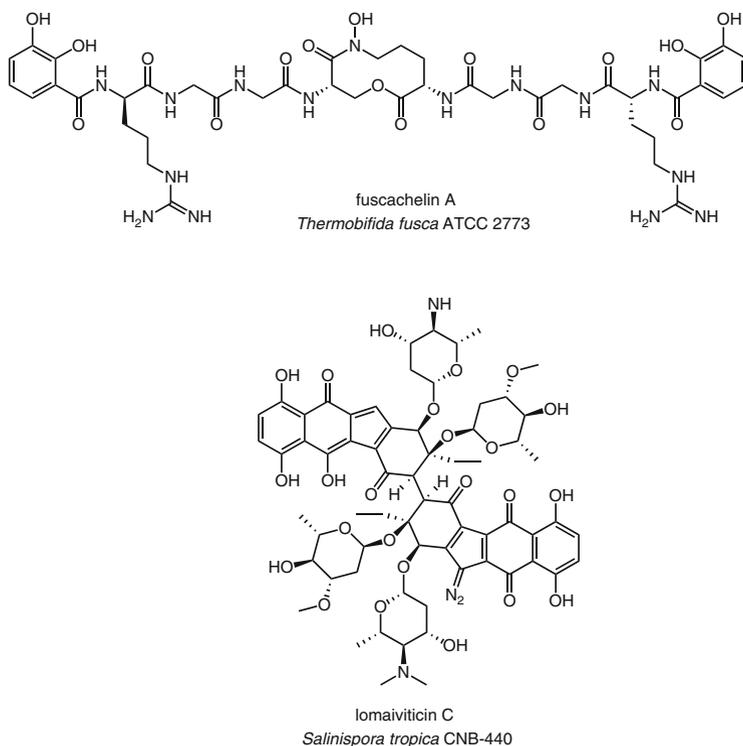
Fig. 11 Examples of natural products identified by their predicted bioactivities

enediyne gene clusters, it is hence possible to link the genetic information with both chemical structures and bioactivity. Those strains that were identified as putative enediyne producers in a genome scanning study were grown in different media, and their culture extracts were analyzed in the biochemical prophage induction assay (BIA) to verify the production of DNA-damaging compounds (*81*, *82*). From the observation that select media supported the production of BIA-active metabolites, the authors concluded that the enediyne biosynthesis genes had been expressed. However, none of the predicted enediynes appears to have been isolated. More recently, a genome mining study, which was conducted in *Salinispora tropica* CNB-440, also revealed the presence of two orphan enediyne pathways. Although testing of *S. tropica* extracts in the BIA indicated the release of DNA-interfering molecules, allelic-exchange mutagenesis and comparative metabolic profiling tracked the observed biological activity to a non-enediyne PKS gene cluster that governed the biosynthesis of lomaiviticin antibiotics (*83*).

Except for the BIA, the only other biological assays, that have been integrated into genome mining studies, are antibacterial assays, which were used to trace lantibiotics from predicted pathways (*84*, *85*). On the other hand, the recent discovery

that microbial gene clusters for the production of proteasome inhibitors include ORFs for specific resistance mechanisms may provide a rationale to utilize proteasome inhibition assays in future genome mining studies (*86*).

4.1.5   One-Strain-Many-Compounds (OSMAC)

One-Strain-Many-Compounds (OSMAC) is a rather undirected, yet promising approach to discover new natural products from microorganisms that is based on the variation of cultivation conditions (*87*). Even though OSMAC studies can be performed independent of the availability of genomic data, the method appears particularly appealing for the analysis of bacteria and fungi, which have previously been identified as potential sources of novel compounds. In an OSMAC study, the fermentation parameters of the analyzed microorganism (media composition, pH, aeration, *etc.*) are successively altered until shifts in the metabolome can be observed (*88*). The OSMAC concept relies on the well-known principle that exogenous stimuli can lead to transcriptional changes, including the activation of otherwise dormant biosynthesis genes.

Several variations of this strategy have been described, including the use of epigenetic modulators (*89*) or the co-cultivation with other microorganisms (*90*). Notwithstanding its empirical approach, OSMAC has been demonstrated repeatedly to be highly useful for retrieving the products of orphan pathways (Fig. 12). Noteworthy examples of natural products that were discovered in this way include jagaricin from the mushroom pathogen *Janthinobacterium agaricidamnosum* (*91*)

jagaricin
Janthinobacterium agaricidamnosum DSM 9628

closthioamide
Clostridium cellulolyticum DSM 5812
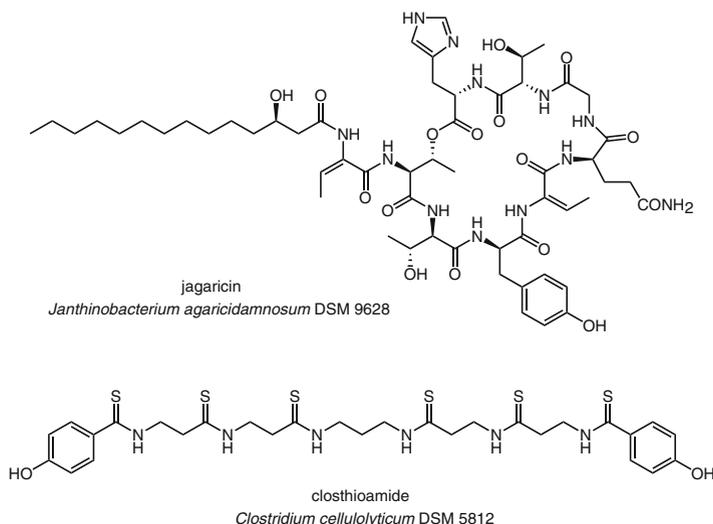
Fig. 12 Secondary metabolites for which their biosynthesis could be induced by mimicking the natural growth conditions of the producing bacteria

and closthioamide, which represents the first antibiotic that was isolated from an anaerobic bacterium (*92*). In both cases, the biosynthesis was triggered upon mimicking environmental growth conditions in the laboratory.

4.2   Genetic Approaches

There are many cases, in which the product of an orphan pathway cannot be predicted accurately. Examples include terpene cyclases, for which the mechanistic logic is still poorly understood (*2*), or PKS- and NRPS-based biosyntheses that deviate from the colinearity rule (*24*, *93*). In other cases, the available genetic information originates from a draft genome sequence, which might not cover a complete biosynthesis locus. Identifying the metabolites associated with these pathways becomes more challenging and often involves genetic engineering to induce, to increase or to abolish their production.

4.2.1   Mutagenesis and Metabolic Profiling

The inactivation of biosynthesis genes and the subsequent chromatographic comparison of mutant and wild types are widely used in natural product discovery (Fig. 13) and pathway assignment. Preferred targets are ORFs, which are considered indispensable for the production of the sought metabolite, because a successful disruption can be expected to completely shut down the respective biosynthesis. In contrast, the inactivation of other genes might have less profound effects.

The first application of the mutagenesis approach in terms of genome mining was described in 2001. Several PKS and NRPS genes were found in the genome of the myxobacterium *Stigmatella aurantiaca*, which could not be correlated with known secondary metabolites. By insertional mutagenesis and comparative metabolic profiling, the pathway-specific products could eventually be identified (*94*). The myxochromides S_{1-3} are cyclic lipopeptides that arise *via* module skipping from a NRPS assembly line (*95*), while the aurafurons are linear polyketides with a terminal furanone ring (*96*). Recently, two gene fragments of *trans*-AT PKSs were targeted in another *S. aurantiaca* strain for disruption. A close examination of extracts from the mutant and wild types revealed the exclusive loss of a single metabolite in the former. This compound was identified as rhizopodin, which was previously not known to be produced by *S. aurantiaca* (*97*). Other genome mining studies in myxobacteria have focused mainly on *Myxococcus xanthus*, which is a model organism for studying fruiting body formation, gliding motility and predation. This soil bacterium had already been established as a prolific source of bioactive secondary metabolites, when the genome of *M. xanthus* DK1622 was sequenced (*98*). The genomic sequence data hence were used primarily to unearth the clusters coding for the biosynthesis of known compounds. Gene inactivation became the preferred method to verify the proposed assignments, not only in DK1622, but also

myxochromide S₁
Stigmatella aurantiaca DW4/3-1

aurafuron A
Stigmatella aurantiaca DW4/3-1

myxoprincomide
Myxococcus xanthus DK1622

coelichelin
Streptomyces coelicolor A3(2)

halstoctacosanolide A
Streptomyces halstedii HC34

bacillaene
Bacillus subtilis

Fig. 13 Examples of secondary metabolites unveiled after gene inactivation and comparative metabolic profiling

in other closely related strains (*99*, *100*). Of note, the genome of *M. xanthus* DK1622 also harbors a large number of orphan biosynthesis pathways. The discovery of myxoprincomide from this strain was a significant achievement that combined the gene inactivation-metabolic profiling approach with an extensive principal-component analysis of preprocessed LC-MS datasets (*101*).

The *Gram*-positive actinomycetes in general and *Streptomyces* spp. in particular are likely the most important microbial resources of clinically used therapeutics. Genome sequencing projects not only confirmed their metabolic prowess, but also

illuminated a huge potential to find novel secondary metabolites (*102*). It is therefore not surprising that actinomycetes have been subjected repeatedly to genome mining. The first study involving a targeted mutagenesis approach was reported from the model strain *Streptomyces coelicolor* A3(2). Its chromosome was found to harbor a gene cluster for a trimodular NRPS, which was predicted to encode the production of a siderophore (*103*). Although a new siderophore named coelichelin was discovered after cultivation in an iron-deficient medium, its structure was inconsistent with the colinearity rule. A gene knockout was hence required to confirm that the targeted NRPS cluster was indeed responsible for coelichelin assembly (*104*). Moreover, deletion of an orphan type III PKS gene in *S. coelicolor* A3(2) and subsequent metabolic analyses revealed the genetic basis for the biosynthesis of the germicidins (*105*). Prior to these studies, two unprecedented 28-membered polyketide macrolactones were reported from *Streptomyces halstedii* HC34. The isolation of the two natural products, which were published with the trivial name halstoctacosanolides, was motivated by the finding of a large type I PKS gene cluster in the genome of *S. halstedii* (*106*). The halstoctacosanolides thus represent the first examples for the genomics-inspired discovery of structurally new natural products from an actinomycete bacterium. Due to violations of the colinearity rule, however, the molecules could not be ascribed unequivocally to the identified gene cluster, and, similar to coelichelin, a gene disruption was ultimately necessary to clarify the assumed correlation (*107*).

A non-canonical pathway was also observed for the antibiotic bacillaene. Originally described in the mid-1990s, the structure and assembly of this protein synthesis inhibitor remained enigmatic until a gene inactivation study established a causal relationship between its production and a cryptic PKS/NRPS locus, which is conserved in the genomes of *B. subtilis* and *B. amyloliquefaciens* FZB 4.2, respectively (*108*, *109*). While previous attempts to structurally characterize bacillaene had been impeded due to its rapid decomposition during isolation, the requirement of a thorough chromatographic purification was circumvented by the generation of a knockout strain. Resonances of the intact natural product were traced subsequently in minimally processed extracts by means of differential NMR analyses (*109*). It should also be noted that the pathways to the known antibiotics difficidin and macrolactin could be identified in *B. amyloliquefaciens* in the course of the bacillaene studies (*108*, *110*).

Aside from myxobacteria, streptomycetes and bacilli, the biosynthesis potential of several other bacteria was tested by targeted mutagenesis (Fig. 14). Interestingly, many of these organisms had not been recognized as natural product sources until their genome sequences became available (*111*). This is most evident in case of β-Proteobacteria belonging to the order *Burkholderiales*, which include some pathogenic members, but also many environmentally and biotechnologically important species. Following the discovery that *Burkholderia rhizoxinica*, an endosymbiont of the plant pathogenic fungus *Rhizopus microsporus*, accounts for the production of the antimitotic agent rhizoxin (*112*), the genus *Burkholderia* came into the spotlight of natural product research. Several new chemical entities have been reported from *Burkholderia* spp. since then, and, in most cases, the identification of these

malleobactin A
Burkholderia thailandensis E264

bactobolin B
Burkholderia thailandensis E264

cupriachelin
Cupriavidus necator H16

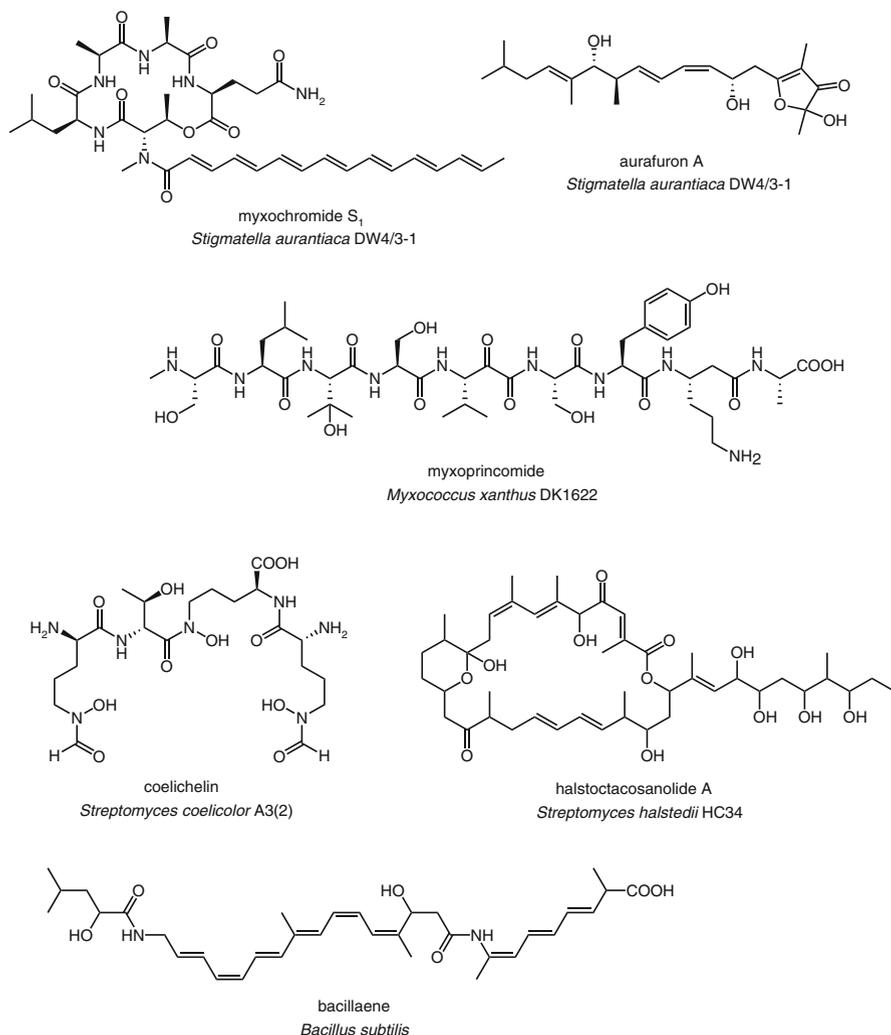delftibactin A
Delftia acidovorans DSM 39

micacocidin
Ralstonia solanacearum GMI1000
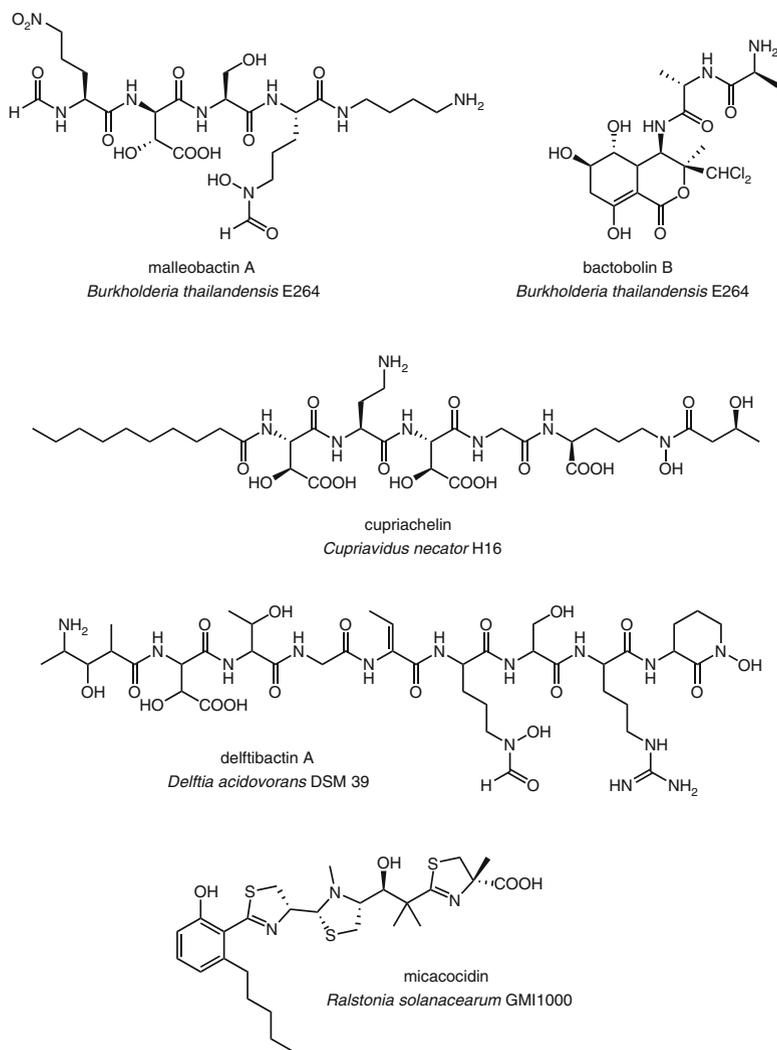
Fig. 14 Examples of secondary metabolites unveiled after gene inactivation and comparative metabolic profiling - continued

compounds involved genome mining techniques. The isolation of the thailandamides has already been mentioned (*70*), but the secondary metabolome of the thailandamide-producing *B. thailandensis* strain E264 turned out to be much more diverse. A subsequent study showed that quorum-sensing signals regulate the production of potent antibiotics, for which the biosynthesis genes could be identified using mutational analysis (*113*). The structures of these antibiotics were eventually resolved and identified as bactobolins (*114*). More recently, the metabolic analysis of a recombinant strain enabled the isolation and structure elucidation of the

long-sought siderophore malleobactin, which might be also produced by human pathogenic *Burkholderia* spp. (*115*). A genomics-driven analysis of the freshwater bacterium *Cupriavidus necator* H16 yielded a series of new siderophores named cupriachelins. These compounds are unusual in that they exhibit structural and physicochemical properties that are usually associated with siderophores from oceanic bacteria, but are only rarely found in other microbial groups, namely, a lipopeptide backbone and photoreactivity (*116*). The bacterium *Delftia acidovorans* can thrive in the presence of toxic gold complexes. Its resistance strategy is based on biomineralization and involves the secretion of a peptidic metallophore, as confirmed by the insertional inactivation of an NRPS gene in the *D. acidovorans* genome. Using a fractionation strategy that was directed toward gold precipitation, the molecule sought could be retrieved in the culture supernatant of the wild type strain. Except for a single event of module skipping, the structure of the isolated gold ligand, delftibactin A, matched the bioinformatic prediction (*117*). A genome mining study in the plant pathogen *Ralstonia solanacearum* GMI1000 revealed a cryptic PKS/NRPS gene cluster, for which the modular assembly line is similar, albeit not identical, to that of the siderophore and virulence factor yersiniabactin. Although yersiniabactin-like biosynthesis clusters were detected in many bacteria (*68*, *118*), only a few of these loci have actually been characterized with respect to their chemical products. A targeted gene disruption confirmed that the *R. solanacearum* cluster is operational under iron-deficient conditions and spurred the isolation of the associated metabolite, which was identified as the antibiotic micacocidin (*119*). The structure of micacocidin was impossible to predict by existing biosynthetic knowledge, and a follow-up study led to the discovery of an unprecedented iteratively acting type I PKS that is integrated into a multimodular assembly (*120*).

The combination of mutagenesis and metabolic profiling has not only been applied in retrieving bacterial natural products. In *Aspergillus nidulans* and *A. terreus*, several orphan gene clusters could be correlated with the production of previously described metabolites by using this strategy (*121–124*).

4.2.2   Transcriptional Modulation

The gene inactivation/metabolic profiling approach, but also the genome mining strategies that exploit structural predictions, presuppose an expression of the biosynthesis genes. A lack of expression means that the compound of interest is not produced and thus cannot be identified. Although it is possible to mediate transcriptional changes by variation of the cultivation conditions, the outcome of such studies is hardly predictable due to their random nature (see Sect. 4.1.5). An alternative option to link cryptic genetic data with secondary metabolites is the reconstitution of the biosynthesis *in vitro* or *in vivo*. However, the applicability of this approach is severely limited in the case of natural product loci, which consist of multiple operons with transcripts in opposite directions or include large biosynthesis genes. With an increasing understanding of transcriptional networks, the manipulation of gene expression has therefore become a favored method to activate silent pathways,

especially in ascomycetes (*125*). Various strategies have been explored in this context that either tend to induce extensive changes in the entire transcriptome of a microorganism or that target the expression of pathway-specific genes. Albeit not necessarily aiming at the activation of a biosynthetic pathway, studies involving the deletion of global regulators will also be covered in this section.

The first study to pursue an expression-directed genome mining approach targeted the nuclear protein LaeA, which can be found in various *Aspergillus* spp. DNA microarray analyses indicated LaeA to act as a pleiotropic regulator of secondary metabolism. Both the deletion and the overexpression of *laeA* in *Aspergillus nidulans* had significant effects on the expression level of several biosynthesis genes and helped to identify those loci that were transcriptionally active. A specific gene disruption in one of these clusters eliminated the production of the antitumor metabolite terrequinone A (Fig. 15), which was previously not known to be produced by the strain analyzed (*126*). A different approach is to target histone and DNA posttranslational processes in order to abrogate the reason for the silencing of certain genes. The repressive chromatin configuration in *A. nidulans* could be relieved by deleting an ORF involved in histone methylation, thereby inducing the biosynthesis of several aromatic polyketides (*127*). Similarly, histone acetylation was modulated. As hypoacetylation of histones is generally correlated with gene silencing, obvious counteractions are to knockout histone deacetylase genes or, conversely, to overexpress histone acetyltransferase genes. Even though either strategy was successfully validated in *A. nidulans*, as evidenced by the increased production levels of known secondary metabolites, no new compounds were reported from these studies (*128*, *129*). In contrast, several previously unrecognized metabolites could be unearthed by manipulating the post-translational modification of non-histone proteins. This is highlighted by the structurally unprecedented pheofungins, which were obtained after the deletion of a putative N-acetyltransferase gene (*130*), as well as by aspernidine A, which was recovered after a kinase knockout in *A. nidulans* (*131*).

The aspyridones were the first natural products for which the discovery involved the transcriptional engineering of pathway-specific genes. It is important to note that the aspyridones are products of a silent PKS/NRPS gene cluster in *A. nidulans*, and that their biosynthesis was not observed until a cluster-associated regulatory gene was homologously overexpressed following a promoter replacement (*132*). Similar strategies were subsequently used for the discovery of the polyketides asperfuranone, neosartoricin, and a dimethylated orthosporin derivative (*133*–*135*). Interestingly, asperfuranone production could also be observed following the induced expression of a regulatory gene belonging to a discrete biosynthesis cluster (*136*). In *A. niger*, the activation of a silent gene cluster triggered the production of naphthacenedione TAN-1612 (*137*). In some cases, in which the biosynthesis involved only a single gene, promoter exchange was also directly applied to the latter, circumventing the need for intervening into higher levels of natural regulation. Examples are csypyrone B1 from *A. oryzae*, which was found after the expression of an orphan type III PKS gene under control of an α-amylase promoter (*138*), and microperfuranone from *A. nidulans* (*139*). Still, a promoter exchange can be associated with certain risks, as indicated by a recent study in *A. terreus*. In this case, the
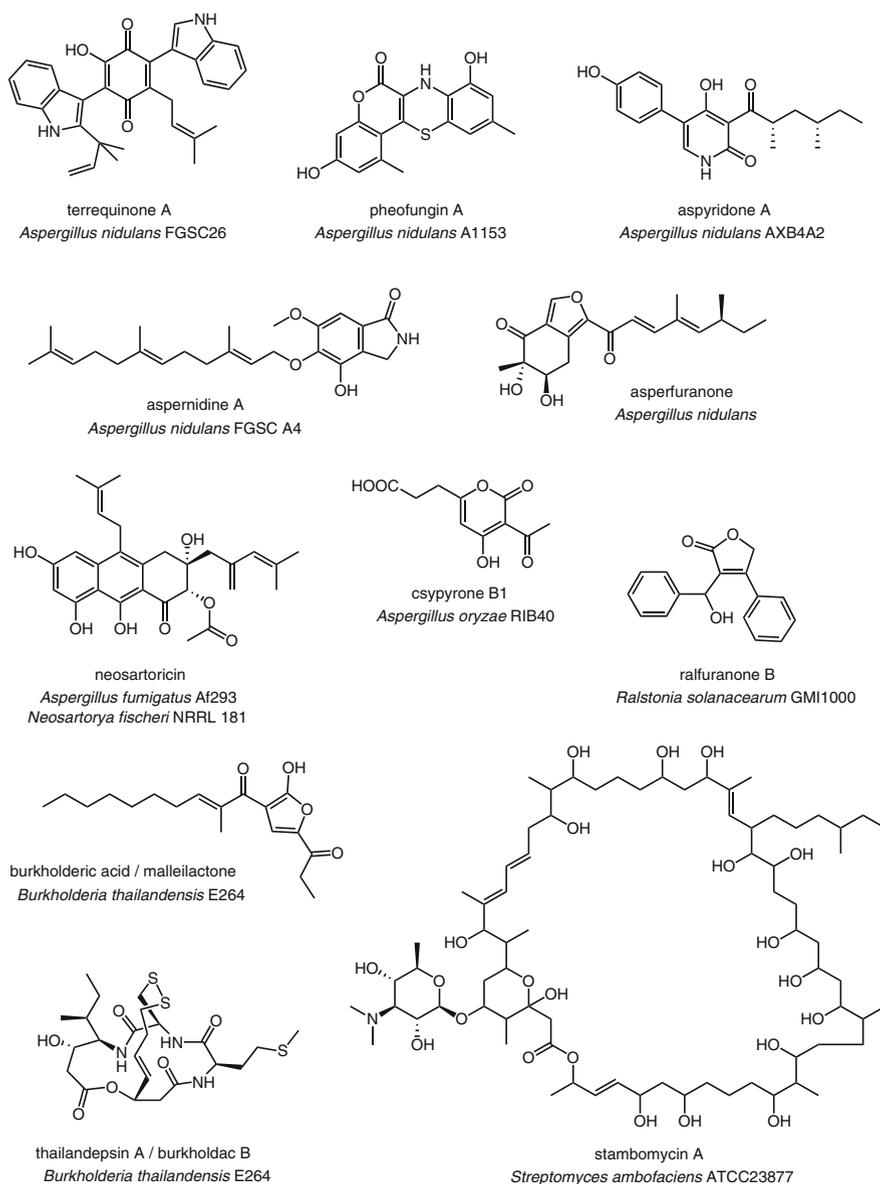
Fig. 15 Natural products identified after transcriptional modulation of the native producer

overexpression of a pathway-specific activator gene under the control of an inducible promoter did not result in product formation, because the inducing agent itself repressed the expression of the biosynthesis genes (*140*).

In the prokaryotic kingdom, transcriptional modulation has been primarily exploited to mine the genomes of *Burkholderiales*. One early report pertains to

R. solanacearum GMI1000, in which crosstalk between global virulence regulators and secondary metabolism promoted the discovery of the ralfuranone family of natural products (*141*). The ralfuranone biosynthesis genes were identified subsequently on the megaplasmid of *R. solanacearum* GMI1000 using a gene inactivation/metabolic profiling approach (*142*). In *B. thailandensis* E264, the inactivation of a LuxR-type gene in the thailandamide locus yielded a previously overlooked lactone derivative (*70, 143*). New depsipeptides, the burkholdacs (syn. thailandepsins), were found following a systematic overexpression of transcription factors associated with secondary metabolite gene clusters (*144*). However, the burkholdacs are not encoded by a silent biosynthesis gene cluster, and they could also be identified by means of a gene deletion approach (*145*). More recently, two groups independently reported on the isolation of a new polyketide natural product named burkholderic acid (syn. malleilactone) from *B. thailandensis* E264. In both studies, a promoter exchange was carried out to trigger the biosynthesis (*146, 147*).

Comparatively few studies were published on the manipulation of regulatory genes in Actinomycetes in order to retrieve the products from their orphan gene clusters. In *Streptomyces ambofaciens*, the in-frame deletion of a pathway-specific repressor led to the continuous production of kinamycin-type antibiotics, which had previously been present only in trace amounts (*148*). Furthermore, the constitutive expression of a regulatory gene within a poorly transcribed PKS cluster of the same strain was key to the discovery of the stambomycins, which represent the largest macrolides ever to be found in an actinomycete. Apart from some unusual structural features, the stambomycins are distinguished by their promising antiproliferative activities (*149*).

4.2.3 Heterologous Expression

If genetic manipulation of the native host is difficult to achieve, the transfer of its silent biosynthesis genes into a suitable host might be the last option for identifying the corresponding natural products (*150*). The same strategy is also indispensable in order to chemically explore metagenomes and environmental DNA assemblies (*151*). Although heterologous expression offers a direct and rational way to compound identification, challenges associated with the assembly of large DNA sequences and, more important, the lack of broadly applicable host systems have as yet limited a wide usage of this powerful technique. Terpene pathways, however, represent an important exception in this context. In one of the first examples of plant genome mining, a cDNA encoding a cryptic oxidosqualene cyclase from *Arabidopsis thaliana* was expressed in yeast to yield the novel triterpene thalianol (Fig. 16) (*152*). This approach circumvented an analysis of the complex metabolite background usually found in plant extracts, and was thus clearly superior to the classical phytochemical methods of natural product discovery. Since then far more triterpenes of *Arabidopsis* have been identified by heterologous expression than by the study of plant material (*153*). Recently, this technique could even be expanded to triterpenoids through the identification of associated enzymes involved in P450 oxidation reactions and subsequent coexpression studies (*154*).
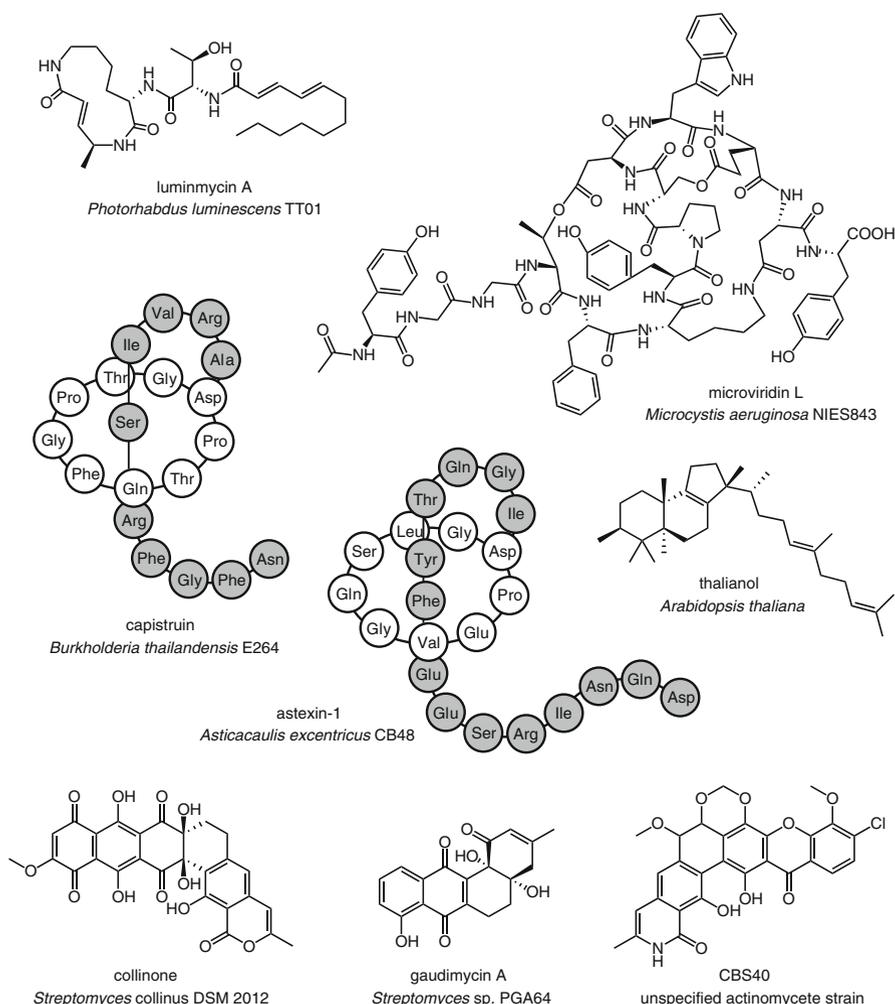
Fig. 16 Secondary metabolites discovered after heterologous expression of their biosynthesis genes

Due to the streamlined genetic organization of their biosynthetic pathways and the restriction to proteinogenic amino acids as building blocks, RiPPs are also attractive candidates for heterologous production (*34*). *Pediococcus pentosaceus* ATCC 25745 belongs to the lactic acid bacteria and is thus a member of a well-known group of bacteriocin producers. As expected, its genome harbors a putative bacteriocin locus, but crucial regulatory and transport genes are absent. This suggests that the gene cluster is not fully functional and, indeed, culture supernatants of *P. pentosaceus* exhibited only negligible antimicrobial activity. The apparently incomplete bacteriocin system was supplemented with the missing ORFs in a heterologous host, allowing for a preliminary characterization of the encoded cryptic

peptide (*155*). A similar strategy was pursued to recover the predicted lasso peptide astexin-1 from the freshwater bacterium *Asticacaulis excentricus*. The lack of a transporter gene, which might confer self-resistance, was held responsible for the absence of astexin-1 in culture extracts of the native producer, and provoked a reconstitution of the biosynthesis in *E. coli* (*156*). The lasso peptide capistruin derived from a non-silent gene cluster of *Burkholderia thailandensis*, but only heterologous expression provided sufficient material for a structural characterization (*157*). The bioinformatic analysis of cyanobacterial genomes suggested a ribosomal biosynthesis for the ecologically significant microviridins. Heterologous expression not only confirmed this assumption (*158*), but it also enabled the identification of new members of this toxin class, including microviridin L (*159*). The sponge-derived polytheonamides are composed of 48 amino acid residues and are thus among the largest peptide natural products known. A recent metagenome mining study revealed that a bacterial endosymbiont is the true source of these impressive compounds (*160*). As expected, the polytheonamides are made *via* a ribosomal pathway. Their biosynthesis requires a total of 48 post-translational modifications, but only six possible candidate enzymes for these reactions could be identified in the polytheonamide cluster. This apparent paucity suggested the repeated use of single enzymes for distinct transformations. The enzymatic balancing act between regiospecificity, on the one hand, and substrate plasticity, on the other, could be confirmed in case of a radical *S*-adenosylmethionine (rSAM) methyltransferase after the coexpression of the corresponding gene with the codon-optimized precursor peptide gene in *E. coli*. Accordingly, the rSAM protein is solely capable to catalyze most of the 18 necessary amino acid epimerizations (*160*).

Only a few studies have reported the successful identification of novel polyketides or nonribosomal peptides by reconstructing their assembly lines in a non-native host. Exceptions are metabolites, for which the biosyntheses are catalyzed by iteratively acting PKSs. Collinone, an angular hexacyclic polyketide, was obtained after its 35-kb gene cluster was transferred from *Streptomyces collinus* DSM 2012 to *S. coelicolor* CH999 (*161*). Further studies involving the interspecies transfer of orphan type II PKS gene loci led to the discovery of the gaudimycins and the chlorinated polyxanthone CBS-40, which possesses promising antibacterial properties (*162*, *163*). Recently, all PKS/NRPS gene clusters from the genome of *Photorhabdus luminescens* TT01 were cloned into *E. coli*-compatible expression vectors. For this purpose, a new strategy was applied, which exploited the increased homologous recombination efficiency of full-length RecE (*164*). Seven out of ten loci were shifted in this way without any mutations, and two of them could be successfully expressed after induction in optimized *E. coli* strains. The observed products included a new glidobactin derivative, luminmycin A, as well as two compounds designated as luminmides. The latter are structurally identical to the GameXPeptides (*74*), which were also discovered from *P. luminescens* with a different approach (see Sect. 4.1.3). *Aspergillus oryzae* served as a host for the reconstitution of pyripyropene biosynthesis. Although pyripyropene is a known natural product, the identity of its putative gene cluster, which was predicted from the genome sequence of *A. fumigatus* Af293, was not clear. The stepwise reconstruction of the meroterpenoid

pathway in *A. oryzae* confirmed the original assumption and further led to the identification of an unprecedented terpene cyclase (*165*).

4.2.4 *In vitro* Reconstitution

For the *in vitro* reconstitution of a biosynthesis pathway, all enzymes involved need to be produced in a soluble, active form. Once available, the respective proteins are incubated together with their predicted substrates as well as cofactors. Owing to the time requirement of this approach and the limited availability of many biosynthetic precursors, *in vitro* reconstitution has been used primarily to investigate the reactions catalyzed by terpene cyclases and lanthionine synthetases (Fig. 17).

epi-isoziaene
Streptomyces coelicolor A3(2)

albaflavenone
Streptomyces coelicolor A3(2)

pentalenene
Streptomyces avermitilis MA-4680

neopentalenolactone D
Streptomyces avermitilis MA-4680

germacradienol
Streptomyces coelicolor A3(2)

geosmin
Streptomyces coelicolor A3(2)

2-methylisoborneol
Streptomyces coelicolor A3(2)

avermitilol
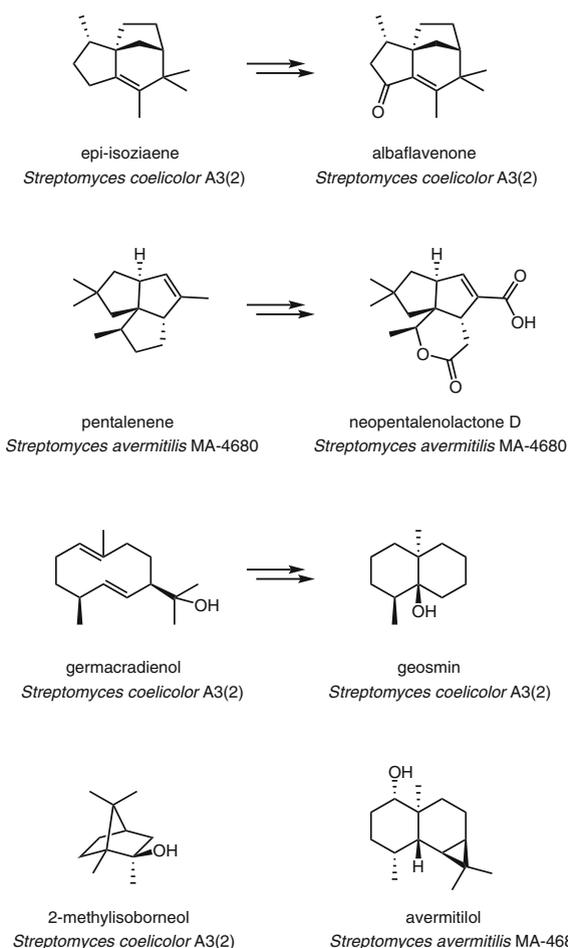*Streptomyces avermitilis* MA-4680

Fig. 17 Natural products for which the biosynthesis pathways were reconstituted *in vitro*

As yet, the low level of sequence similarity among terpene cyclases has precluded the development of reliable models to predict their biochemical function, whereas the experimental analysis of recombinant proteins is well established (*41*). However, care must be taken in interpreting the results of *in vitro* assays with single enzymes. The identified products can be the true end products of a metabolic pathway but, more often, they turn out to be biosynthetic intermediates or, in few cases, even shunt products. Examples include the identification of the epi-isozizaene synthase and the pentalenene synthase in *Streptomyces* spp. (*166*, *167*). In the former case, a cytochrome P450 gene is located adjacent to the terpene synthase gene. An overlap of the two transcripts already suggested a functional correlation and, indeed, the recombinant CYP450 mediated two consecutive allylic oxidations in epi-isozizaene, culminating in albaflavenone formation (*168*). In case of the pentalenene synthase, the corresponding gene was found to reside in a cluster comprising at least 13 protein coding sequences. This locus was expected to have a role in pentalenolactone biosynthesis, but a successive reconstitution revealed a pathway to the unprecedented neopentalenolactones. The production of these pentalenolactone isomers was ultimately confirmed by chemical analyses (*169*). SCO6073, a cryptic sesquiterpene synthase from *Streptomyces coelicolor* A3(2), had been initially correlated with germacradienol biosynthesis based upon *in vitro* data (*170*). Later, it was proposed that the assembly of germacradienol is just the first reaction in a multistep pathway to the earthy odorant geosmin (*171*). While it was speculated originally that geosmin biosynthesis involves the action of several enzymes in addition to SCO6073, further investigations confirmed that the latter is solely capable of converting farnesyl pyrophosphate into geosmin with germacradienol as an intermediate (*172*). The pathway to another odorous terpene, 2-methylisoborneol, from *S. coelicolor* was also reconstituted *in vitro*. This led to the unexpected finding that the required C-methylation precedes the formation of the bicyclo[2.2.1]heptanol ring (*173*, *174*). The novel sesquiterpene alcohol avermitilol was discovered after incubation of an orphan terpene synthase from *S. avermitilis* with farnesyl pyrophosphate (*175*). Following its homologous overexpression in a genome-minimized mutant, the production of avermitilol could also be verified *in vivo* (*176*).

Haloduracin was the first RiPP that was discovered by applying a genome mining strategy and for which the biosynthesis was reconstituted *in vitro* after the production of the predicted two-component lantibiotic had been confirmed in *Bacillus halodurans* C-125 by means of bioactivity testing (see 4.1.4) and MS analyses (*84*). Notably, the dehydration sites and the ring topology of the haloduracin constituents were not determined from the isolated *Bacillus*-derived haloduracin, but from intermediates that had been generated in reactions of purified, recombinant enzymes. In lantibiotic biosynthesis, the final maturation, *i.e.* the removal of an N-terminal leader sequence, typically is carried out by a membrane-bound protease. Since the expression of this type of enzyme is challenging, the natural cleavage sites in the haloduracin components were replaced with an engineered Factor Xa recognition sequence. This elegant approach allowed the usage of a commercially available protease, thereby concluding the pathway reconstitution (*84*). An interesting example of combinatorial RiPP biosynthesis was found in the cyanobacterium

Prochlorococcus marinus MIT9313. *In vitro* testing of a single lanthionine synthetase unveiled its astonishing capability to convert up to 29 different precursor peptides into an assembly of natural products with diverse ring topologies. The genes encoding the precursor peptides were scattered throughout the *Prochlorococcus* chromosome, yet the production of the anticipated lantipeptides could be confirmed in late-exponential cultures of the native host (*177*). A new type of lanthionine synthetase that consists of three distinct catalytic domains was identified in the translated genome sequence of *Streptomyces venezuelae* ATCC10712. Reconstitution of its enzyme activity using individually expressed and purified domains revealed that the thioether ring formation involves the transient phosphorylation of serine and threonine residues in the precursor peptide (*178*).

5   Conclusions and Perspectives

After about a decade of applied genome mining, it is evident that significant progress has been made in harnessing the idle biosynthetic potential of natural product-producing organisms. This development was clearly favored by the implementation of improved analytical techniques in the genome mining process (*75*, *101*), as well as by the utilization of genetic tools to induce the expression of otherwise silent biosynthesis genes. The modulation of chromatin-mediated gene regulation (*127*) and the engineering of ribosomal mutations in bacteria (*179*, *180*) are only two among the many, new exciting methods that have become available. Another noteworthy contribution was the introduction of user-friendly bioinformatic software to support the prediction of chemical structures from genomic data (*63*).

The prerequisite for successful genome mining is a sound understanding of the genetic and enzymatic basis of natural product biosynthesis, the foundations for which had already been laid through pioneering work in the 1990s. It is important to note that the knowledge gained in the field of biosynthesis has not ceased. Instead, it can be expected that new insights will further improve the translation of genetic into chemical information. The large number of post-translationally modified ribosomal peptides that have been identified recently, is exemplary in this context (*34*). On the other hand, fundamental aspects on thiotemplate-based assembly systems are, contrary to the general opinion, still far from being understood. The fragmentary knowledge about domain interactions as well as iterative processes not only affects the reliability of bioinformatic predictions, but also a general ability to manipulate and reprogram the corresponding enzymes. These unsolved issues must be addressed in the upcoming years. Another crucial task will be the development of appropriate genetic systems for such organisms that are as yet not amenable to recombinant engineering (*181*). The realization that many biosynthesis genes fail to give rise to products in the laboratory unless they are specifically activated underpins the importance of such efforts. An obvious alternative to the genetic manipulation of the native producer is heterologous expression. Supposedly, this approach will to play a more dominant role in the genomics-based prospecting of

microorganisms and plants for new drugs in the future. Although much development work still lies ahead, important milestones have already been achieved with the validation of robust methods for the cloning and/or assembly of large DNA sequences (*164*, *182*).

References

1. Lerner CG, Hajduk PJ, Wagner R, Wagenaar FL, Woodall C, Gu Y-G, Searle XB, Florjancic AS, Zhang T, Clark RF, Cooper CS, Mack JC, Yu L, Cai M, Betz SF, Chovan LE, McCall JO, Black-Schaefer CL, Kakavas SJ, Schurdak ME, Comess KM, Walter KA, Edalji R, Dorwin SA, Smith RA, Hebert EJ, Harlan JE, Metzger RE, Merta PJ, Baranowski JL, Coen ML, Thornewell SJ, Shivakumar AG, Saiki AY, Soni N, Bui M, Balli DJ, Sanders WJ, Nilius AM, Holzman TF, Fesik SW, Beutel BA (2007) From Bacterial Genomes to Novel Antibacterial Agents: Discovery, Characterization, and Antibacterial Activity of Compounds that Bind to HI0065 (YjeE) from *Haemophilus influenzae*. Chem Biol Drug Des 69:395
2. Challis GL (2008) Genome Mining for Novel Natural Product Discovery. J Med Chem 51:2618
3. Gross H (2009) Genomic Mining – a Concept for the Discovery of New Bioactive Natural Products. Curr Opin Drug Discov Develop 12:207
4. Wecke T, Mascher T (2011) Antibiotic Research in the Age of Omics: from Expression Profiles to Interspecies Communication. J Antimicrob Chemother 66:2689
5. Donadio S, Monciardini P, Sosio M (2007) Polyketide Synthases and Nonribosomal Peptide Synthetases: the Emerging View from Bacterial Genomics. Nat Prod Rep 24:1073
6. Gross H (2007) Strategies to Unravel the Function of Orphan Biosynthesis Pathways: Recent Examples and Future Prospects. Appl Microbiol Biotechnol 75:267
7. Cragg GM, Newman DJ (2013) Natural Products: A Continuing Source of Novel Drug Leads. Biochim Biophys Acta 1830:3670
8. Baltz RH (2008) Renaissance in Antibacterial Discovery from Actinomycetes. Curr Opin Pharmacol 8:557
9. Gourdeau H, McAlpine JB, Ranger M, Simard B, Berger, Beaudry F, Farnet CM, Falardeau P (2008) Identification, Characterization and Potent Antitumor Activity of ECO-4601, a Novel Peripheral Benzodiazepine Receptor Ligand. Cancer Chemorther Pharmacol 61:911
10. Miao V, Coëffet-LeGal M-F, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH (2005) Daptomycin Biosynthesis in *Streptomyces roseosporus*: Cloning and Analysis of the Gene Cluster and Revision of Peptide Stereochemistry. Microbiology 151:1507
11. Regalado A (1999) Mining the Genome. Technol Rev 102:56
12. The Arabidopsis Genome Initiative (2000) Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. Nature 408:796
13. Salcedo T, Upadhyay RJ, Nagasaki K, Bhattacharya D (2012) Dozens of Toxin-related Genes are Expressed in a Nontoxic Strain of the Dinoflagellate *Heterocapsa circularisquama*. Mol Biol Evol 29:1503
14. Kliebenstein DJ, Osbourn A (2012) Making New Molecules – Evolution of Pathways for Novel Metabolites in Plants. Curr Opin Plant Biol 15:415

15. Martin JF (1992) Clusters of Genes for the Biosynthesis of Antibiotics: Regulatory Genes and Overproduction of Pharmaceuticals. J Ind Microbiol 9:73
16. Koonin EV (2009) Evolution of Genome Architecture. Int J Biochem Cell Biol 41:298
17. Hertweck C (2009) The Biosynthetic Logic of Polyketide Diversity. Angew Chem Int Ed 48:4688
18. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF (1990) An Unusually Large Multifunctional Polypeptide in the Erythromycin-producing Polyketide Synthase of *Saccharopolyspora erythraea*. Nature 348:176
19. Reeves CD, Murli S, Ashley GW, Piagentini M, Hutchinson CR, McDaniel R (2001) Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-specific Mutations. Biochemistry 40:15464
20. Price MN, Huang KH, Alm EJ, Arkin AP (2005) A Novel Method for Accurate Operon Predictions in all Sequenced Prokaryotes. Nucleic Acids Res 33:880
21. Ishikawa J, Hotta K (1999) FramePlot: a New Implementation of the Frame Analysis for Predicting Protein-coding Regions in Bacterial DNA with a High G+C Content. FEMS Microbiol Lett 174:251
22. Wilson MC, Moore BS (2012) Beyond Ethylmalonyl-CoA: The Functional Role of Crotonyl-CoA Carboxylase/Reductase Homologs in Expanding Polyketide Diversity. Nat Prod Rep 29:72
23. Udwary DW, Gontang EA, Jones AC, Jones CS, Schultz AW, Winter JM, Yang JY, Beauchemin N, Capson TL, Clark BR, Esquenazi E, Eustáquio AS, Freel K, Gerwick L, Gerwick WH, Gonzalez D, Liu W-T, Malloy KL, Maloney KN, Nett M, Nunnery JK, Penn K, Prieto-Davo A, Simmons TL, Weitz S, Wilson MC, Tisa LS, Dorrestein PC, Moore BS (2011) Significant Natural Product Biosynthetic Potential of Actinorhizal Symbionts of the Genus *Frankia*, as Revealed by Comparative Genomic and Proteomic Analyses. Appl Environ Microbiol 77:3617
24. Piel J (2010) Biosynthesis of Polyketides by *trans*-AT Polyketide Synthases. Nat Prod Rep 27:996
25. Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A (2007) Type II Polyketide Synthases: Gaining a Deeper Insight into Enzymatic Teamwork. Nat Prod Rep 24:162
26. Austin MB, Noel JP (2003) The Chalcone Synthase Superfamily of Type III Polyketide Synthases. Nat Prod Rep 20:79
27. Marahiel MA (2009) Working Outside the Protein-synthesis Rules: Insights into Non-ribosomal Peptide Synthesis. J Pept Sci 15:799
28. MacCabe AP, van Liempt H, Palissa H, Unkles SE, Riach MBR, Pfeifer E, von Döhren H, Kinghorn JR (1991) δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine Synthetase from *Aspergillus nidulans*. J Biol Chem 266:12646
29. Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural Basis for the Activation of Phenylalanine in the Nonribosomal Biosynthesis of Gramicidin S. EMBO J 16:4174
30. Stachelhaus T, Mootz HD, Marahiel MA (1999) The Specificity-conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. Chem Biol 6:493
31. Challis GL, Ravel J, Townsend CA (2000) Predictive, Structure-based Model of Amino Acid Recognition by Nonribosomal Peptide Synthetase Adenylation Domains. Chem Biol 7:211
32. Schwecke T, Göttling K, Durek P, Dueñas I, Käufer NF, Zock-Emmenthal S, Staub E, Neuhof T, Dieckmann R, von Döhren H (2006) Nonribosomal Peptide Synthesis in *Schizosaccharomyces pombe* and the Architectures of Ferrichrome-type Siderophore Synthetases in Fungi. ChemBioChem 7:612
33. Fischbach MA, Walsh CT (2006) Assembly-line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. Chem Rev 106:3468
34. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schnidt EW, Selsted ME, Severinov K, Shen B, Sivonen

K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. Nat Prod Rep 30:108

35. Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (2005) Patellamide A and C Biosynthesis by a Microcin-like Pathway in *Prochloron didemni*, the Cyanobacterial Symbiont of *Lissoclinum patella*. Proc Natl Acad Sci USA 102:7315

36. Zhang Q, Yu Y, Vélasquez JE, van der Donk WA (2012) Evolution of Lanthipeptide Synthetases. Proc Natl Acad Sci USA 109:18361

37. Marsh AJ, O'Sullivan O, Ross RP, Cotter PD, Hill C (2010) *In silico* Analysis Highlights the Frequency and Diversity of Type I Lantibiotic Gene Clusters in Genome Sequenced Bacteria. BMC Genomics 11:679

38. Daum M, Herrmann S, Wilkinson B, Bechthold A (2009) Genes and Enzymes Involved in Bacterial Isoprenoid Biosynthesis. Curr Opin Chem Biol 13:180

39. Oldfield E, Lin F-Y (2012) Terpene Biosynthesis: Modularity Rules. Angew Chem Int Ed 51:1124

40. Dairi T (2005) Studies on Biosynthetic Genes and Enzymes of Isoprenoids Produced by Actinomycetes. J Antibiot 58:227

41. Cane DE, Ikeda H (2012) Exploration and Mining of the Bacterial Terpenome. Acc Chem Res 45:463

42. Jenke-Kodama H, Dittmann E (2009) Bioinformatic Perspectives on NRPS/PKS Megasynthases: Advances and Challenges. Nat Prod Rep 26:874

43. Yadav G, Gokhale RS, Mohanty D (2003) SEARCHPKS: a Program for Detection and Analysis of Polyketide Synthase Domains. Nucleic Acids Res 31:3654

44. Ansari MZ, Yadav G, Gokhale RS, Mohanty D (2004) NRPS-PKS: a Knowledge-based Resource for Analysis of NRPS/PKS Megasynthases. Nucleic Acids Res 32:W405

45. Yadav G, Gokhale RS, Mohanty D (2003) Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. J Mol Biol 328:335

46. Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, Ansari MZ, Mohanty D (2010) SBSPKS: Structure Based Sequence Analysis of Polyketide Synthases. Nucleic Acids Res 38:W487

47. Tae H, Kong E-B, Park K (2007) ASMPKS: an Analysis System for Modular Polyketide Synthases. BMC Bioinformatics 8:327

48. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) *ClustScan*: an Integrated Program Package for the Semi-automatic Annotation of Modular Biosynthetic Gene Clusters and *In Silico* Prediction of Novel Chemical Structures. Nucleic Acids Res 36:6882

49. Li MHT, Ung PMU, Zajkowski J, Garneau-Tsodikova S, Sherman DH (2009) Automated Genome Mining for Natural Products. BMC Bioinformatics 10:185

50. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH (2005) Specificity Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Transductive Support Vector Machines (TSVMs). Nucleic Acids Res 33:5799

51. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O (2011) NRPSpredictor2 - a Web Server for Predicting NRPS Adenylation Domain Specificity. Nucleic Acids Res 39:W362

52. Prieto C, Garcia-Estrada C, Lorenzana D, Martín JF (2012) NRPSsp: Non-ribosomal Peptide Synthase Substrate Predictor. Bioinformatics 28:426

53. Khayatt BI, Overmars L, Siezen RJ, Francke C (2013) Classification of the Adenylation and Acyl-transferase Activity of NRPS and PKS Systems Using Ensembles of Substrate Specific Hidden *Markov* Models. PLoS ONE 8:e62136

54. Minowa Y, Araki M, Kanehisa M (2007) Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. J Mol Biol 368:1500

55. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The Natural Product Domain Seeker NaPDoS: a Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. PLoS ONE 7:e34064

56. Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, Wohlleben W (2009) CLUSEAN: A Computer-based Framework for the Automated Analysis of Bacterial Secondary Metabolite Gene Clusters. J Biotechnol 140:13

57. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: Genomic Mapping of Fungal Secondary Metabolite Clusters. Fungal Genet Biol 47:736

58. Kim J, Yi G-S (2012) PKMiner: a Database for Exploring Type II Polyketide Synthases. BMC Microbiol 12:169

59. Kamra P, Gokhale RS, Mohanty D (2005) SEARCHGTr: a Program for Analysis of Glycosyltransferases Involved in Glycosylation of Secondary Metabolites. Nucleic Acids Res 33:W220

60. de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP (2006) BAGEL: a Web-based Bacteriocin Genome Mining Tool. Nucleic Acids Res 34:W273

61. de Jong A, van Heel AJ, Kok J, Kuipers OP (2010) BAGEL2: Mining for Bacteriocins in Genomic Data. Nucleic Acids Res 38:W647

62. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. Nucleic Acids Res 39:W339

63. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013) antiSMASH 2.0 - a Versatile Platform for Genome Mining of Secondary Metabolite Producers. Nucleic Acids Res 41:W204

64. Woodward RB (1942) Structure and Absorption Spectra. III. Normal Conjugated Dienes. J Am Chem Soc 64:72

65. McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco A-M, Zazopoulos E, Farnet CM (2005) Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation: ECO-02301, a Novel Antifungal Agent, as an Example. J Nat Prod 68:493

66. Banskota AH, McAlpine JB, Sorensen D, Ibrahim A, Aouidate M, Piraee M, Alarco A-M, Farnet CM, Zazopoulos E (2006) Genomic Analyses Lead to Novel Secondary Metabolites. Part 3 ECO-0501, a Novel Antibacterial of a New Class. J Antibiot 59:533

67. Banskota AH, McAlpine JB, Sorensen D, Aouidate M, Piraee M, Alarco A-M, Omura S, Shiomi K, Farnet CM, Zazopoulos E (2006) Isolation and Identification of Three New 5-Alkenyl-3,3(2*H*)-furanones from Two *Streptomyces* Species Using a Genomic Screening Approach. J Antibiot 59:168

68. Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS (2007) Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete *Salinispora tropica*. Proc Natl Acad Sci USA 104:10376

69. Nett M, Moore BS (2009) Exploration and Engineering of Biosynthetic Pathways in the Marine Actinomycete *Salinispora tropica*. Pure Appl Chem 81:1075

70. Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J (2008) Exploiting the Mosaic Structure of *trans*-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. Nat Biotechnol 26:225

71. Scherlach K, Hertweck C (2006) Discovery of Aspoquinolones A-D, Prenylated Quinoline-2-one Alkaloids from *Aspergillus nidulans*, Motivated by Genome Mining. Org Biomol Chem 4:3517

72. Gross H, Stockwell VO, Henkels MD, Nowak-Thompson B, Loper JE, Gerwick WH (2007) The Genomisotopic Approach: a Systematic Method to Isolate the Products of Orphan Biosynthetic Gene Clusters. Chem Biol 14:53

73. Kreutzer MF, Nett M (2012) Genomics-driven Discovery of Taiwachelin, a Lipopeptide Siderophore from *Cupriavidus taiwanensis*. Org Biomol Chem 10:9338

74. Bode HB, Reimer D, Fuchs SW, Kirchner F, Dauth C, Kegler C, Lorenzen W, Brachmann AO, Grün P (2012) Determination of the Absolute Configuration of Peptide Natural Products by Using Stable Isotope Labeling and Mass Spectrometry. Chem Eur J 18:2342

75. Kersten RD, Yang Y-L, Xu Y, Cimermancic P, Nam S-J, Fenical W, Fischbach MA, Moore BS, Dorrestein PC (2011) A Mass Spectrometry-guided Genome Mining Approach for Natural Product Peptidogenomics. Nat Chem Biol 7:794

76. Liu W-T, Kersten RD, Yang Y-L, Moore BS, Dorrestein PC (2011) Imaging Mass Spectrometry and Genome Mining *via* Short Sequence Tagging Identified the Anti-infective Agent Arylomycin in *Streptomyces roseosporus*. J Am Chem Soc 133:18010

77. Fuchs SW, Sachs CC, Kegler C, Nollmann FI, Karas M, Bode HB (2012) Neutral Loss Fragmentation Pattern Based Screening for Arginine-rich Natural Products in *Xenorhabdus* and *Photorhabdus*. Anal Chem 84:6948

78. Schwyn B, Neilands JB (1987) Universal Chemical Assay for the Detection and Determination of Siderophores. Anal Biochem 160:47

79. Dimise EJ, Widboom PF, Bruner SD (2008) Structure Elucidation and Biosynthesis of Fuscachelins, Peptide Siderophores from the Moderate Thermophile *Thermobifida fusca*. Proc Natl Acad Sci USA 105:15311

80. Liu W, Ahlert J, Gao Q, Wendt-Pienkowski E, Shen B, Thorson JS (2003) Rapid PCR Amplification of Minimal Enediyne Polyketide Synthase Cassettes Leads to a Predictive Familial Classification Model. Proc Natl Acad Sci USA 100:11959

81. Zazopoulos E, Huang K, Staffa A, Liu W, Bachmann BO, Nonaka K, Ahlert J, Thorson JS, Shen B, Farnet CM (2003) A Genomics-guided Approach for Discovering and Expressing Cryptic Metabolic Pathways. Nat Biotechnol 21:187

82. Elespuru RK, Yarmolinsky MB (1979) A Colorimetric Assay of Lysogenic Induction Designed for Screening Potential Carcinogenic and Carcinostatic Agents. Environ Mutagen 1:65

83. Kersten RD, Lane AL, Nett M, Richter TKS, Duggan BM, Dorrestein PC, Moore BS (2013) Bioactivity-guided Genome Mining Reveals the Lomaiviticin Biosynthetic Gene Cluster in *Salinispora tropica*. ChemBioChem 14:955

84. McClerren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA (2006) Discovery and *in Vitro* Biosynthesis of Haloduracin, a Two-component Lantibiotic. Proc Natl Acad Sci USA 103:17243

85. Begley M, Cotter PD, Hill C, Ross RP (2009) Identification of a Novel Two-peptide Lantibiotic, Lichenicidin, Following Rational Genome Mining for LanM Proteins. Appl Environ Microbiol 75:5451

86. Kale AJ, McGlinchey RP, Lechner A, Moore BS (2011) Bacterial Self-resistance to the Natural Proteasome Inhibitor Salinosporamide A. ACS Chem Biol 6:1257

87. Bode HB, Bethe B, Höfs R, Zeeck A (2002) Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. ChemBioChem 3:619

88. Nett M, Hertweck C (2011) Farinamycin, a Quinazoline from *Streptomyces griseus*. J Nat Prod 74:2265

89. Cichewicz RH (2010) Epigenome Manipulation as a Pathway to New Natural Product Scaffolds and Their Congeners. Nat Prod Rep 27:11

90. Schroekh V, Scherlach K, Nützmann H-W, Shelest E, Schmidt-Heck W, Schuemann J, Martin K, Hertweck C, Brakhage AA (2009) Intimate Bacterial-Fungal Interaction Triggers Biosynthesis of Archetypal Polyketides in *Aspergillus nidulans*. Proc Natl Acad Sci USA 106:14558

91. Graupner K, Scherlach K, Bretschneider T, Lackner G, Roth M, Gross H, Hertweck C (2012) Imaging Mass Spectrometry and Genome Mining Reveal Highly Antifungal Virulence Factor of Mushroom Soft Rot Pathogen. Angew Chem Int Ed 51:13173

92. Lincke T, Behnken S, Ishida K, Roth M, Hertweck C (2010) Closthioamide: an Unprecedented Polythioamide Antibiotic from the Strictly Anaerobic Bacterium *Clostridium cellulolyticum*. Angew Chem Int Ed 49:2011

93. Wenzel SC, Müller R (2005) Formation of Novel Secondary Metabolites by Bacterial Multimodular Assembly Lines: Deviations from Textbook Biosynthetic Logic. Curr Opin Chem Biol 9:447

94. Silakowski B, Kunze B, Müller R (2001) Multiple Hybrid Polyketide Synthase/Non-ribosomal Peptide Synthetase Gene Clusters in the Myxobacterium *Stigmatella aurantiaca*. Gene 275:233

95. Wenzel SC, Kunze B, Höfle G, Silakowski B, Scharfe M, Blöcker H, Müller R (2005) Structure and Biosynthesis of Myxochromides S$_{1-3}$ in *Stigmatella aurantiaca*: Evidence for

an Iterative Bacterial Type I Polyketide Synthase and for Module Skipping in Nonribosomal Peptide Biosynthesis. ChemBioChem 6:375

96. Kunze B, Reichenbach H, Müller R, Höfle G (2005) Aurafuron A and B, New Bioactive Polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (Myxobacteria). J Antibiot 58:244

97. Pistorius D, Müller R (2012) Discovery of the Rhizopodin Biosynthetic Gene Cluster in *Stigmatella aurantiaca* Sg a15 by Genome Mining. ChemBioChem 13:416

98. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB (2006) Evolution of Sensory Complexity Recorded in a Myxobacterial Genome. Proc Natl Acad Sci USA 103:15200

99. Wenzel SC, Müller R (2009) The Impact of Genomics on the Exploitation of the Myxobacterial Secondary Metabolome. Nat Prod Rep 26:1385

100. Cortina NS, Revermann O, Krug D, Müller R (2011) Identification and Characterization of the Althiomycin Biosynthetic Gene Cluster in *Myxococcus xanthus* DK897. ChemBioChem 12:1411

101. Cortina NS, Krug D, Plaza A, Revermann O, Müller R (2012) Myxoprincomide: a Natural Product from *Myxococcus xanthus* Discovered by Comprehensive Analysis of the Secondary Metabolome. Angew Chem Int Ed 51:811

102. Nett M, Ikeda H, Moore BS (2009) Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. Nat Prod Rep 26:1362

103. Challis GL, Ravel J (2000) Coelichelin, a New Peptide Siderophore Encoded by the *Streptomyces coelicolor* Genome: Structure Prediction from the Sequence of its Nonribosomal Peptide Synthetase. FEMS Microbiol Lett 187:111

104. Lautru S, Deeth RJ, Bailey LM, Challis GL (2005) Discovery of a New Peptide Natural Product by *Streptomyces coelicolor* Genome Mining. Nat Chem Biol 1:265

105. Song L, Barona-Gomez F, Corre C, Xiang L, Udwary DW, Austin MB, Noel JP, Moore BS, Challis GL (2006) Type III Polyketide Synthase β-Ketoacyl-ACP Starter Unit and Ethylmalonyl-CoA Extender Unit Selectivity Discovered by *Streptomyces coelicolor* Genome Mining. J Am Chem Soc 128:14754

106. Tohyama S, Eguchi T, Dhakal RP, Akashi T, Otsuka M, Kakinuma K (2004) Genome-inspired Search for New Antibiotics. Isolation and Structure Determination of New 28-Membered Polyketide Macrolactones, Halstoctacosanolides A and B, from *Streptomyces halstedii* HC34. Tetrahedron 60:3999

107. Tohyama S, Kakinuma K, Eguchi T (2006) The Complete Biosynthetic Gene Cluster of the 28-Membered Polyketide Macrolactones, Halstoctacosanolides, from *Streptomyces halstedii* HC34. J Antibiot 59:44

108. Chen X-H, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süssmuth RD, Borriss R (2006) Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in *Bacillus amyloliquefaciens* FZB42. J Bacteriol 188:4024

109. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J (2007) The Identification of Bacillaene, the Product of the PksX Megacomplex in *Bacillus subtilis*. Proc Natl Acad Sci USA 104:1506

110. Schneider K, Chen X-H, Vater J, Franke P, Nicholson G, Borriss R, Süssmuth RD (2007) Macrolactin is the Polyketide Biosynthesis Product of the *pks2* Cluster of *Bacillus amyloliquefaciens* FZB42. J Nat Prod 70:1417

111. Winter JM, Behnken S, Hertweck C (2011) Genomics-inspired Discovery of Natural Products. Curr Opin Chem Biol 15:22

112. Partida-Martinez LP, Hertweck C (2005) Pathogenic Fungus Harbours Endosymbiotic Bacteria for Toxin Production. Nature 437:884

113. Duerkop BA, Varga J, Chandler JR, Peterson SB, Herman JP, Churchill ME, Parsek MR, Nierman WC, Greenberg EP (2009) Quorum Sensing Control of Antibiotic Synthesis in *Burkholderia thailandensis*. J Bacteriol 191:3909

114. Seyedsayamdost MR, Chandler JR, Blodgett JAV, Lima PS, Duerkop BA, Oinuma K-I, Greenberg EP, Clardy J (2010) Quorum-Sensing-Regulated Bactobolin Production by *Burkholderia thailandensis* E264. Org Lett 12:716

115. Franke J, Ishida K, Ishida-Ito M, Hertweck C (2013) Nitro Versus Hydroxamate in Siderophores of Pathogenic Bacteria: Effect of Missing Hydroxylamine Protection in Malleobactin Biosynthesis. Angew Chem Int Ed 52:8271

116. Kreutzer MF, Kage H, Nett M (2012) Structure and Biosynthetic Assembly of Cupriachelin, a Photoreactive Siderophore from the Bioplastic Producer *Cupriavidus necator* H16. J Am Chem Soc 134:5415

117. Johnston CW, Wyatt MA, Li X, Ibrahim A, Shuster J, Southam G, Magarvey NA (2013) Gold Biomineralization by a Metallophore from a Gold-Associated Microbe. Nat Chem Biol 9:241

118. Waterfield NR, Sanchez-Contreras M, Eleftherianos I, Dowling A, Yang G, Wilkinson P, Parkhill J, Thomson N, Reynolds SE, Bode HB, Dorus S, ffrench-Constant RH (2008) Rapid Virulence Annotation (RVA): Identification of Virulence Factors Using a Bacterial Genome Library and Multiple Invertebrate Hosts. Proc Natl Acad Sci USA 105:15967

119. Kreutzer MF, Kage H, Gebhardt P, Wackler B, Saluz HP, Hoffmeister D, Nett M (2011) Biosynthesis of a Complex Yersiniabactin-like Natural Product *via* the *mic* Locus in Phytopathogen *Ralstonia solanacearum*. Appl Environ Microbiol 77:6117

120. Kage H, Kreutzer MF, Wackler B, Hoffmeister D, Nett M (2013) An Iterative Type I Polyketide Synthase Initiates the Biosynthesis of the Antimycoplasma Agent Micacocidin. Chem Biol 20:764

121. Chiang YM, Szewczyk E, Nayak T, Davidson AD, Sanchez JF, Lo HC, Ho WY Simityan H, Kuo E, Praseuth A, Watanabe K, Oakley BR, Wang CC (2008) Molecular Genetic Mining of the *Aspergillus* Secondary Metabolome: Discovery of the Emericellamide Biosynthetic Pathway. Chem Biol 15:527

122. Sanchez JF, Entwistle R, Hung J-H, Yaegashi J, Jain S, Chiang Y-M, Wang CCC, Oakley BR (2011) Genome-based Deletion Analysis Reveals the Prenyl Xanthone Biosynthesis Pathway in *Aspergillus nidulans*. J Am Chem Soc 133:4010

123. Lo H-C, Entwistle R, Guo C-J, Ahuja M, Szewczyk E, Hung J-H, Chiang Y-M, Oakley BR, Wang CCC (2012) Two Separate Gene Clusters Encode the Biosynthetic Pathway for the Meroterpenoids Austinol and Dehydroaustinol in *Aspergillus nidulans*. J Am Chem Soc 134:4709

124. Guo C-J, Yeh H-H, Chiang Y-M, Sanchez JF, Chang S-L, Bruno KS, Wang CCC (2013) Biosynthetic Pathway for the Epipolythiodioxopiperazine Acetylaranotin in *Aspergillus terreus* Revealed by Genome-based Deletion Analysis. J Am Chem Soc 135:7205

125. Brakhage AA (2013) Regulation of Fungal Secondary Metabolism. Nat Rev Microbiol 11:21

126. Bok JW, Hoffmeister D, Maggio-Hall LA, Renato M, Glasner JD, Keller NP (2006) Genomic Mining for *Aspergillus* Natural Products. Chem Biol 13:31

127. Bok JW, Chiang Y-M, Szewczyk E, Reyes-Dominguez Y, Davidson AD, Sanchez JF, Lo H-C, Watanabe K, Strauss J, Oakley BR, Wang CCC, Keller NP (2009) Chromatin-level Regulation of Biosynthetic Gene Clusters. Nat Chem Biol 5:462

128. Shwab EK, Bok JW, Tribus M, Galehr J, Graessle S, Keller NP (2007) Histone Deacetylase Activity Regulates Chemical Diversity in *Aspergillus*. Eukaryot Cell 6:1656

129. Soukup AA, Chiang Y-M, Bok JW, Reyes-Dominguez Y, Oakley BR, Wang CCC, Strauss J, Keller NP (2012) Overexpression of the *Aspergillus nidulans* Histone 4 Acetyltransferase EsaA Increases Activation of Secondary Metabolite Production. Mol Microbiol 86:314

130. Scherlach K, Nützmann H-W, Schroeckh V, Dahse H-M, Brakhage AA, Hertweck C (2011) Cytotoxic Pheofungins from an Engineered Fungus Impaired in Posttranslational Protein Modification. Angew Chem Int Ed 50:9843

131. Yaegashi J, Praseuth MB, Tyan S-W, Sanchez JF, Entwistle R, Chiang Y-M, Oakley BR, Wang CCC (2013) Molecular Genetic Characterization of the Biosynthesis Cluster of a Prenylated Isoindolinone Alkaloid Aspernidine A in *Aspergillus nidulans*. Org Lett 15:2862

132. Bergmann S, Schümann J, Scherlach K, Lange C, Brakhage AA, Hertweck C (2007) Genomics-driven Discovery of PKS-NRPS Hybrid Metabolites from *Aspergillus nidulans*. Nat Chem Biol 3:213

133. Chiang YM, Szewczyk E, Davidson AD, Keller N, Oakley BR, Wang CC (2009) A Gene Cluster Containing Two Fungal Polyketide Synthases Encodes the Biosynthetic Pathway for a Polyketide, Asperfuranone, in *Aspergillus nidulans*. J Am Chem Soc 131:2965

134. Chooi Y-H, Fang J, Liu H, Filler SG, Wang P, Tang Y (2013) Genome Mining of a Prenylated and Immunosuppressive Polyketide from Pathogenic Fungi. Org Lett 15:780

135. Nakazawa T, Ishiuchi K, Praseuth A, Noguchi H, Hotta K, Watanabe K (2012) Overexpressing Transcriptional Regulator in *Aspergillus oryzae* Activates a Silent Biosynthetic Pathway to Produce a Novel Polyketide. ChemBioChem 13:855

136. Bergmann S, Funk AN, Scherlach K, Schroeckh V, Shelest E, Horn U, Hertweck C, Brakhage AA (2010) Activation of a Silent Fungal Polyketide Biosynthesis Pathway Through Regulatory Cross Talk with a Cryptic Nonribosomal Peptide Synthetase Gene Cluster. Appl Environ Microbiol 76:8143

137. Li Y, Chooi Y-H, Sheng Y, Valentine JS, Tang Y (2011) Comparative Characterization of Fungal Anthracenone and Naphthacenedione Biosynthetic Pathways Reveals an α-Hydroxylation-dependent *Claisen*-like Cyclization Catalyzed by a Dimanganese Thioesterase. J Am Chem Soc 133:15773

138. Seshime Y, Juvvadi PR, Kitamoto K, Ebizuka, Fujii I (2010) Identification of Csypyrone B1 as the Novel Product of *Aspergillus oryzae* Type III Polyketide Synthase CsyB. Bioorg Med Chem 18:4542

139. Yeh H-H, Chiang Y-M, Entwistle R, Ahuja M, Lee K-H, Bruno KS, Wu T-K, Oakley BR, Wang CCC (2012) Molecular Genetic Analysis Reveals that a Nonribosomal Peptide Synthetase-like (NRPS-like) Gene in *Aspergillus nidulans* is Responsible for Microperfuranone Biosynthesis. Appl Microbiol Biotechnol 96:739

140. Gressler M, Zaehle C, Scherlach K, Hertweck C, Brock M (2011) Multifactorial Induction of an Orphan PKS-NRPS Gene Cluster in *Aspergillus terreus*. Chem Biol 18:198

141. Schneider P, Jacobs JM, Neres J, Aldrich CC, Allen C, Nett M, Hoffmeister D (2009) The Global Virulence Regulators VsrAD and PhcA Control Secondary Metabolism in the Plant Pathogen *Ralstonia solanacearum*. ChemBioChem 10:2730

142. Wackler B, Schneider P, Jacobs JM, Pauly J, Allen C, Nett M, Hoffmeister D (2011) Ralfuranone Biosynthesis in *Ralstonia solanacearum* Suggests Functional Divergence in the Quinone Synthetase Family of Enzymes. Chem Biol 18:354

143. Ishida K, Lincke T, Behnken S, Hertweck C (2010) Induced Biosynthesis of Cryptic Polyketide Metabolites in a *Burkholderia thailandensis* Quorum Sensing Mutant. J Am Chem Soc 132:13966

144. Biggins JB, Gleber CD, Brady SF (2011) Acyldepsipeptide HDAC Inhibitor Production Induced in *Burkholderia thailandensis*. Org Lett 13:1536

145. Wang C, Henkes LM, Doughty LB, He M, Wang D, Meyer-Almes F-J, Cheng Y-Q (2011) Thailandepsins: Bacterial Products with Potent Histone Deacetylase Inhibitory Activities and Broad-spectrum Antiproliferative Activities. J Nat Prod 74:2031

146. Franke J, Ishida K, Hertweck C (2012) Genomics-driven Discovery of Burkholderic Acid, a Noncanonical, Cryptic Polyketide from Human Pathogenic *Burkholderia* Species. Angew Chem Int Ed 51:11611

147. Biggins JB, Ternei MA, Brady SF (2012) Malleilactone, a Polyketide Synthase-Derived Virulence Factor Encoded by the Cryptic Secondary Metabolome of *Burkholderia pseudomallei* Group Pathogens. J Am Chem Soc 134:13192

148. Bunet R, Song L, Mendes MV, Corre C, Hotel L, Rouhier N, Framboisier X, Leblond P, Challis GL, Aigle B (2011) Characterization and Manipulation of the Pathway-Specific Late Regulator AlpW Reveals *Streptomyces ambofaciens* as a New Producer of Kinamycins. J Bacteriol 193:1142

149. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B (2011) Identification of a Bioactive 51-Membered Macrolide Complex by Activation of a Silent Polyketide Synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci USA 108:6258

150. Zotchev SB, Sekurova ON, Katz L (2012) Genome-based Bioprospecting of Microbes for New Therapeutics. Curr Opin Biotechnol 23:941
151. Brady SF, Simmons L, Kim JH, Schmidt EW (2009) Metagenomic Approaches to Natural Products from Free-living and Symbiotic Organisms. Nat Prod Rep 26:1488
152. Fazio GC, Xu R, Matsuda SPT (2004) Genome Mining to Identify New Plant Triterpenoids. J Am Chem Soc 126:5678
153. Morlacchi P, Wilson WK, Xiong Q, Bhaduri A, Sttivend D, Kolesnikova MD, Matsuda SPT (2009) Product Profile of PEN3: the Last Unexamined Oxidosqualene Cyclase in *Arabidopsis thaliana*. Org Lett 11:2627
154. Castillo DA, Kolesnikova MD, Matsuda SPT (2013) An Effective Strategy for Exploring Unknown Metabolic Pathways by Genome Mining. J Am Chem Soc 135:5885
155. Diep DB, Godager L, Brede D, Nes IF (2006) Data Mining and Characterization of a Novel Pediocin-like Bacteriocin System from the Genome of *Pediococcus pentosaceus* ATCC 25745. Microbiology 152:1649
156. Maksimov MO, Pelczer I, Link AJ (2012) Precursor-Centric Genome-Mining Approach for Lasso Peptide Discovery. Proc Natl Acad Sci USA 109:15223
157. Knappe TA, Linne U, Zirah S, Rebuffat S, Xie X, Marahiel MA (2008) Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of *Burkholderia thailandensis* E264. J Am Chem Soc 130:11446
158. Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E (2008) Ribosomal Synthesis of Tricyclic Depsipeptides in Bloom-forming Cyanobacteria. Angew Chem Int Ed 47:7756
159. Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E (2010) Exploiting the Natural Diversity of Microviridin Gene Clusters for Discovery of Novel Tricyclic Depsipeptides. Appl Environ Microbiol 76:3568
160. Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, Sahl H-G, Matsunaga S, Piel J (2012) Metagenome Mining Reveals Polytheonamides as Posttranslationally Modified Ribosomal Peptides. Science 338:387
161. Martin R, Sterner O, Alvarez MA, De Clercq E, Bailey JE, Minas W (2001) Collinone, a New Recombinant Angular Polyketide Antibiotic Made by an Engineered *Streptomyces* Strain. J Antibiot 54:239
162. Palmu K, Ishida K, Mäntsälä P, Hertweck C, Metsä-Ketelä M (2007) Artificial Reconstruction of Two Cryptic Angucycline Antibiotic Biosynthetic Pathways. ChemBioChem 8:1577
163. Hornung A, Bertazzo M, Dziarnowski A, Schneider K, Welzel K, Wohlert SE, Holzenkämpfer M, Nicholson GJ, Bechthold A, Süssmuth RD, Vente A, Pelzer S (2007) A Genomic Screening Approach to the Structure-Guided Identification of Drug Candidates from Natural Sources. ChemBioChem 8:757
164. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Müller R, Stewart AF, Zhang Y (2012) Full-length RecE Enhances Linear-Linear Homologous Recombination and Facilitates Direct Cloning for Bioprospecting. Nat Biotechnol 30:440
165. Itoh T, Tokunaga K, Matsuda Y, Fujii I, Abe I, Ebizuka Y, Kushiro T (2010) Reconstitution of a Fungal Meroterpenoid Biosynthesis Reveals the Involvement of a Novel Family of Terpene Cyclases. Nat Chem 2:858
166. Lin X, Hopson R, Cane DE (2006) Genome Mining in *Streptomyces coelicolor*: Molecular Cloning and Characterization of a New Sesquiterpene Synthase. J Am Chem Soc 128:6022
167. Tetzlaff CN, You Z, Cane DE, Takamatsu S, Omura S, Ikeda H (2006) A Gene Cluster for Biosynthesis of the Sesquiterpenoid Antibiotic Pentalenolactone in *Streptomyces avermitilis*. Biochemistry 45:6179
168. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, Cane DE (2008) Biosynthesis of the Sesquiterpene Antibiotic Albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem 283:8183
169. Jiang J, Tetzlaff CN, Takamatsu S, Iwatsuki M, Komatsu M, Ikeda H, Cane DE (2009) Genome Mining in *Streptomyces avermitilis*: A Biochemical *Baeyer-Villiger* Reaction and Discovery of a New Branch of the Pentalenolactone Family Tree. Biochemistry 48:6431

170. Cane DE, Watt RM (2003) Expression and Mechanistic Analysis of a Germacradienol Synthase from *Streptomyces coelicolor* Implicated in Geosmin Biosynthesis. Proc Natl Acad Sci USA 100:1547

171. Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted *Streptomyces* Gene Replacement Identifies a Protein Domain Needed for Biosynthesis of the Sesquiterpene Soil Odor Geosmin. Proc Natl Acad Sci USA 100:1541

172. Jiang J, He X, Cane DE (2007) Biosynthesis of the Earthy Odorant Geosmin by a Bifunctional *Streptomyces coelicolor* Enzyme. Nat Chem Biol 3:711

173. Komatsu M, Tsuda M, Omura S, Oikawa H, Ikeda H (2008) Identification and Functional Analysis of Genes Controlling Biosynthesis of 2-Methylisoborneol. Proc Natl Acad Sci USA 105:7422

174. Wang CM, Cane DE (2008) Biochemistry and Molecular Genetics of the Biosynthesis of the Earthy Odorant Methylisoborneol in *Streptomyces coelicolor*. J Am Chem Soc 130:8908

175. Chou WK, Fanizza I, Uchiyama T, Komatsu M, Ikeda H, Cane DE (2010) Genome Mining in *Streptomyces avermitilis*: Cloning and Characterization of SAV_76, the Synthase for a New Sesquiterpene, Avermitilol. J Am Chem Soc 132:8850

176. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H (2010) Genome-Minimized *Streptomyces* Host for the Heterologous Expression of Secondary Metabolism. Proc Natl Acad Sci USA 107:2646

177. Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA (2010) Catalytic Promiscuity in the Biosynthesis of Cyclic Secondary Metabolites in Planktonic Marine Cyanobacteria. Proc Natl Acad Sci USA 107:10430

178. Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA (2010) Discovery of Unique Lanthionine Synthetases Reveals New Mechanistic and Evolutionary Insights. PLoS Biol 8:e1000339

179. Hosaka T, Ohnishi-Kameyama M, Muramatsu H, Murakami K, Tsurumi Y, Kodani S, Yoshida M, Fujie A, Ochi K (2009) Antibacterial Discovery in Actinomycetes Strains with Mutations in RNA Polymerase or Ribosomal Protein S12. Nat Biotechnol 27:462

180. Gomez-Escribano JP, Song L, Fox DJ, Yeo V, Bibb MJ, Challis GL (2012) Structure and Biosynthesis of the Unusual Polyketide Alkaloid Coelimycin P1, a Metabolic Product of the *cpk* Gene Cluster of *Streptomyces coelicolor* M145. Chem Sci 3:2716

181. Scharf DH, Brakhage AA (2013) Engineering Fungal Secondary Metabolism: A Roadmap to Novel Compounds. J Biotechnol 163:179

182. Ongley SE, Bian X, Neilan BA, Müller R (2013) Recent Advances in the Heterologous Expression of Microbial Natural Product Biosynthetic Pathways. Nat Prod Rep 30:1121