Handbook of
NATURALLY OCCURRING COMPOUNDS

Volume I
Acetogenins, Shikimates, and Carbohydrates

T. K. DEVON

A. I. SCOTT

Handbook of
NATURALLY OCCURRING COMPOUNDS

Volume I
Acetogenins, Shikimates, and Carbohydrates

T. K. DEVON

PFIZER CENTRAL RESEARCH LABORATORIES
SANDWICH, ENGLAND

A. I. SCOTT

STERLING CHEMISTRY LABORATORY
YALE UNIVERSITY, NEW HAVEN, CONNECTICUT

Academic Press Inc., New York / San Francisco London 1975
A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Devon, T K
Handbook of naturally occurring compounds.

CONTENTS: v. 1. Acetogenins, shikimates, and carbohydrates. v. 2. Terpenes.
1. Natural products–Handbooks, manuals, etc.
I. Scott, Alastair L., joint author. II. Title.
III. Title: Naturally occurring compounds.
QD415.7.D48 547'.7'0212 76-187258
ISBN 0-12-213601-2 (v. 1)

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

With the advent of spectroscopy as a tool in the structural elucidation of natural products, there has been a rapid increase in our knowledge of the structures and variety of naturally occurring compounds. It has consequently become an increasingly difficult task to maintain current awareness in the natural products field, and the acquisition of retrospective data is a problem only partly alleviated at present by literature reviews, monographs, and compendia. It was in an attempt to pool the chemical and biochemical data of natural products that the Card Index File of Naturally Occurring Compounds was conceived and initiated at the University of Sussex in 1966. This project, the first essential step in the construction of a more comprehensive system, undertook to search for, list, and classify all reported naturally occurring compounds the structures of which had been determined (with the exclusion of polymeric and macromolecular compounds). The interest so often expressed in this file led, in 1970, to the reorganization of part of the data into the publishable format now presented.

It is somewhat inevitable, when attempting so wide-ranging a project, that the selection criteria should be controversial; thus, the decision to include steroidal aglycones as "natural products" is likely to be frowned upon by the purists. Others, however, may consider the absence of antibiotic degradation products a severe limitation. Similarly, the lack of inclusion, at this stage, of spectroscopic and botanical data may appear limiting; however, an attempt to be so comprehensive so rapidly would have greatly delayed the availability of the present material. It was felt that the present format and material could well stand alone in satisfying particular needs in this field; the provision of a literature reference for each compound should assist in the further acquisition of data.

No pretence of complete comprehensiveness is being claimed for this handbook, nor will it likely be found to be error free; however, considerable care has been taken to abstract the literature as deeply and thoroughly as our resources have permitted. A certain degree of editing of material for structural and sterochemical correlations has been undertaken and has hopefully removed the bulk of errors and ambiguities from the handbook. It would, of course, be greatly appreciated if diligent users who spot errors or omissions would forward them to the authors so that subsequent yearbooks can set the records straight.

We would like to indicate our debt to the numerous authors of earlier compendia, monographs, and reviews from whose works the original Card Index File of Naturally Occurring Compounds gained much of its foundation. For the facilities enabling the construction of the File our appreciation is tendered to Chemical Laboratory, University of Sussex. At Yale University we gratefully received assistance and encouragement from many colleagues in the Sterling Chemistry Laboratory. Dr. Paul Reichardt (University of Alaska) has become a valuable participant in this undertaking, and we look forward to his active collaboration with us in the production of the supplementary volumes. The final preparation

of material for publication was ably performed by Mrs. Diane Devon with contributions to the indexing from Jeffery and Ronald Burns. Finally our appreciation is extended to Pfizer Central Research for the facilities and atmosphere which enable this endeavor to continue to flourish.

GUIDE TO HANDBOOK USAGE

Contents

This Handbook contains at present most of the known naturally occurring compounds to which structures have been assigned.

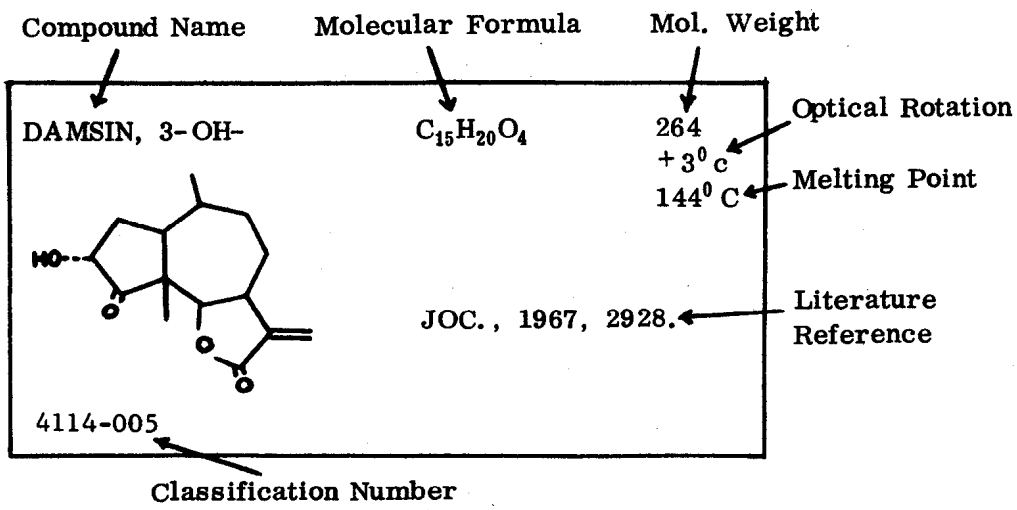
Limitations

Excluded from the Handbook are polymeric compounds, such as proteins and polysaccharides, synthetic derivatives of natural products, and degradation products (artifacts).

Data

Each structure is stored in the Handbook with its name, molecular formula, molecular weight, optical rotation (α_D), melting point, literature reference (usually the latest), and classification number.

DATA FORMAT



Indices

Compounds can be retrieved by name using the ALPHABETICAL INDEX or by structural type using the STRUCTURAL CLASSIFICATION GUIDES at the beginning of each class. MOLECULAR WEIGHT AND MOLECULAR FORMULA INDICES are also provided at the end of the book.

Code Numbers

CLASSIFICATION CODE NUMBERS are used to specify structural types, and the key to these is held within the individual Structural Guides. The first two digits represent the major class and the second two the subclass. Each individual compound also possesses a COMPOUND SEQUENCE NUMBER which, in combination with the Classification Code Number, supplies a unique address for that compound. The compounds are stored in the Handbook in ascending sequence, new compounds being simply inserted at the end of its appropriate section.

Classification Guide

The Catalogue of each primary class of compounds will be preceded by its corresponding Classification Guide which in general consists of:

Introduction
Biogenetic Chart
Main Skeleton Index
Less Common Skeletons Index

Main Volumes

The wide range of structural types of naturally occurring compounds have been classified and then collected into three groups, each of which will be issued as a separate volume:

Volume I	Acetogenins, Shikimates, and Carbohydrates
Volume II	Terpenes
Volume III	Alkaloids and Related Nitrogenous Compounds

Supplements

Each April an annual supplement for the preceding year will be published for each volume which will contain new compounds, structure changes in earlier reported compounds, and any additional data and indices. At suitable intervals these supplements will be cumulated and merged into the main volumes.

ABBREVIATIONS

Nonstandard Journal Abbreviations

Aust. J. C.	Australian Journal of Chemistry
Can. J. C.	Canadian Journal of Chemistry
Helv.	Helvetica Chimica Acta
Ind. J. C.	Indian Journal of Chemistry
JACS	Journal of the American Chemical Society
JCS	Journal of the Chemical Society
JOC	Journal of Organic Chemistry
J. Ind. C.S.	Journal of the Indian Chemical Society

Symbols Used for Solvents (optical rotations)

a	Acetone	m	Methanol
b	Benzene	n	HCl
c	Chloroform	p	Pyridine
d	Dioxan	r	Diethyl ether
e	Ethanol	t	Carbon tetrachloride
h	Hexane	w	Water

Other Abbreviations

ac	Acetate	Glu	Glucosyl
Ac	Acetyl	Mann	Mannosyl
AMB	Alpha-methyl-butyryl	Me	Methyl
Amorph	Amorphous	m.e.	Methyl ester
Ang	Angeloyl	OI	Optically Inactive
Ara	Arabinosyl	Ph	Phenyl
Bu	Butyl	pic	Picrate
Bz	Benzoyl	Pr	Propyl
Caff	Caffeoyl	Rac	Racemic
Cinn	Cinnamoyl	Rha	Rhamnosyl
Coum	Coumaroyl	Sen	Senecioyl
DMA	Dimethylallyl	Tig	Tigloyl
Et	Ethyl	Val	Valeryl
Fruct	Fructosyl	Van	Vanilloyl
Ger	Geranyl	Ver	Veratroyl

CONTENTS

INTRODUCTION

Of the three volumes that constitute this handbook, this particular compilation presents the greatest challenge to introduction. It would be a simple matter to state the contents of the other volumes and contend that this volume provides coverage of all other classes of natural products. Alas, such an abdication cannot be. In biogenetic terms the compounds included in this volume are derived from one (or more) of four major pathways:

shikimic acid pathway
acetate/malonate pathway
carbohydrate metabolism
tricarboxylic acid cycle

Thus, Volume I coverage can be clearly differentiated from that of Volume II (terpenes derived from the mevalonate pathway). Volume III material (alkaloids, amino acids, etc.) cannot be so readily distinguished in these terms, for the aromatic amino acids are products of the shikimic acid pathway, and some classes of alkaloids are of acetogenic origin. Suffice it to say here that the majority of nitrogenous compounds are located in Volume III, notable exceptions being the tetracyclines and aminosaccharides, which have been placed in Volume I, and the terpene alkaloids located in Volume II. This volume also has the distinction of embracing the largest collection of compounds, some 4800 of the 12,000 natural products in our files (as of early 1972). These structurally defined compounds have been assigned a two-digit Primary Classification Code Number(PCCN) that ranges from 01 to 35 plus a "miscellaneous" class, PCCN = 39, which has been provided for complex or ambiguous structures. The primary classification is given in the table overleaf, secondary subclassifications being outlined at the beginning of the appropriate primary sections of the handbook. Because of some of the classification problems associated with compounds such as the phenols, some effort has been taken to provide introductions to the biogenetic schemes utilized in this volume. It should be borne in mind that such structural/biogenetic schemes are frequently hypothetical simplifications, and the inclusion of a compound in a particular class is in no way a comment on its actual biosynthetic origin. Thus, while the anthraquinone class is considered to be a member of the acetogenic group, it is now well known that several plant anthraquinones are derived via the shikimic acid pathway.

THE PRIMARY CLASSES IN VOLUME I

Shikimate Aromatics

01 C_6-C_1
02 C_6-C_2
03 C_6-C_3
04 C_6-C_n ($n \geqslant 4$)
05 Coumarins
06 Lignans
07 Terphenyls
19 Prearomatics

Acetate/Shikimate Aromatics

09 Neoflavonoids
10 C_6-C_3-C_6
11 Flavones
12 Flavanones
13 Flavans
14 Isoflavonoids
15 Biflavonoids
16 Pterocarpans/Coumestans
17 Anthocyanins
18 Rotenoids/Peltogynans/ Aurones
28 Benzophenones/Stilbenes

Acetate/Malonate Aromatics

21 Orcinols
22 Phloroglucinols
23 Depsides/Depsidones
24 Xanthones
25 Naphthoquinones
26 Anthraquinones
27 Polycyclics
28 Biphenyls/Grisans

Acyclic and Heterocyclic Acetogenins

31 Hydrocarbons, Fatty Acids, etc.
32 Oxyheterocyclics and Phenylpolyynes
33 Thio Compounds
34 Macrolides

Carbohydrates

35 Saccharides/Cyclitols/ Aminoglycosides

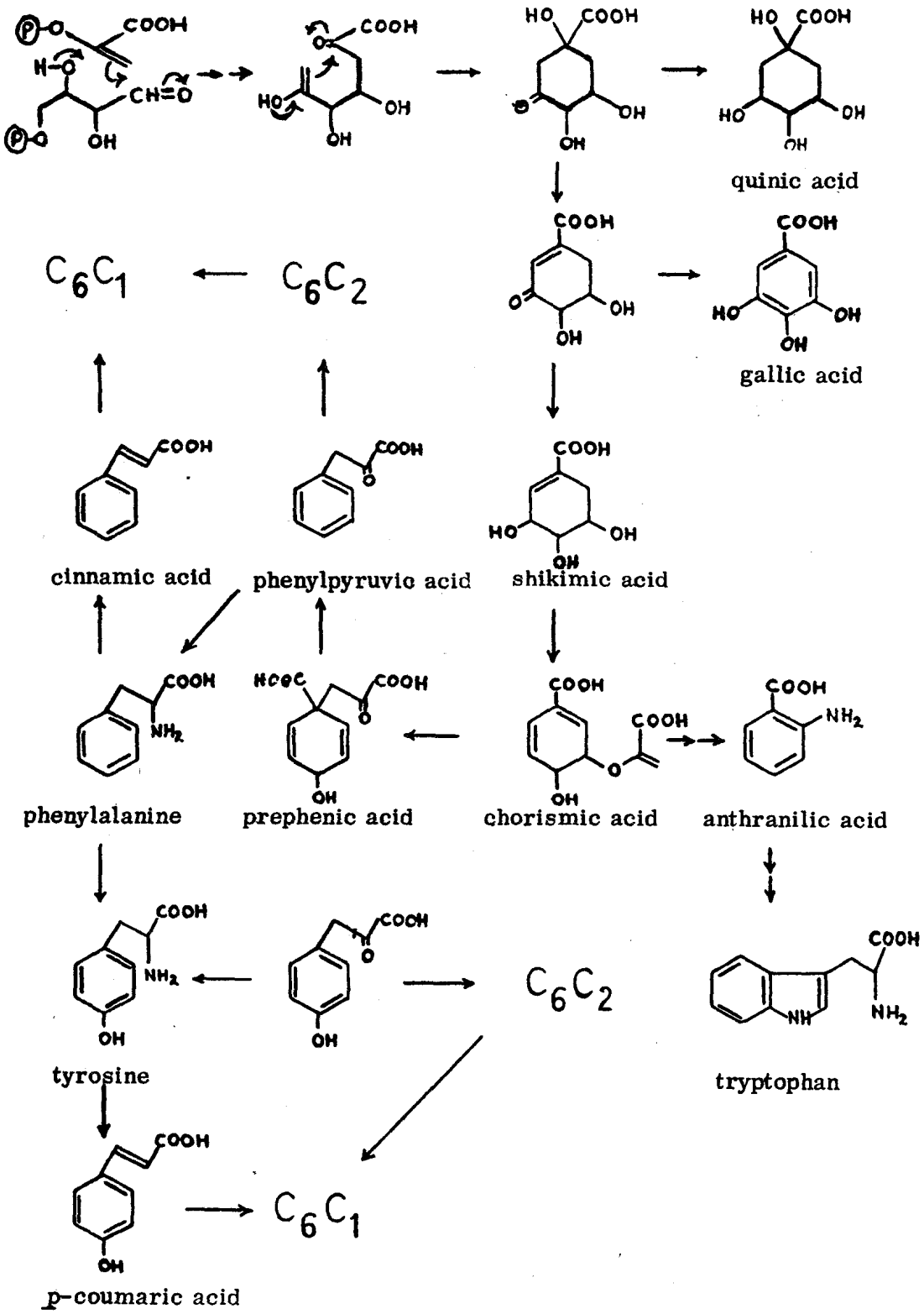
Complex Classes

08 Phenols (not classified elsewhere)
19 Phenol Epoxides/ Shikimate Prearomatics
20 Benzoquinones
28 C_6-C_n-C_6
29 Cyclopentanoids
30 Tetronic Acids/ Tropolones/ Pyrones/ Furans
39 Miscellaneous

SHIKIMATE-DERIVED AROMATICS

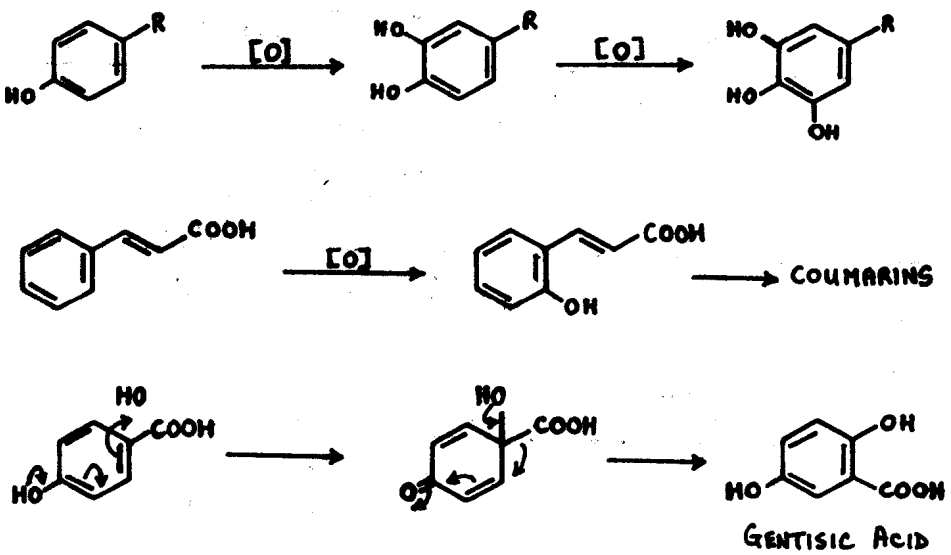
Aromatic compounds of shikimic acid origin in general are characterized by their oxygenation patterns, lack of nuclear alkylation, and limited structural modification. The scheme provided on the next page summarizes the shikimic acid pathway and the relationships between the aromatics derived via shikimic acid. From this scheme it should be evident how the structural characteristics of shikimates originate and serve to distinguish them from acetogenic aromatics. It is of course important to realize that reliance on such features for biogenetic analysis is a double-edged tool and can lead to "misplacing" shikimate-looking aromatics of acetogenic origin. The overall utility of the classification does, however, significantly outweigh such reservations.

Chart 1: The Shikimic Acid Pathway

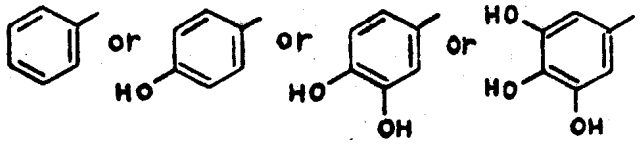


Oxygenation Patterns of Shikimate-Derived Aromatics

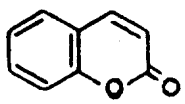
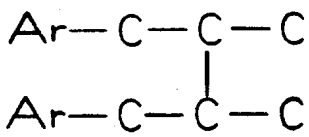
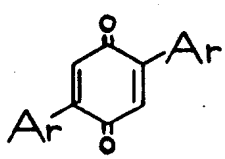
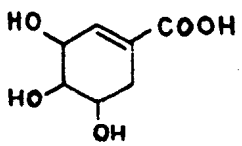
As has already been intimated, aromatic compounds derived from shikimic acid frequently display characteristic oxygenation patterns. The origin of these patterns can be rationalized quite simply by a consideration of the biogenetic scheme illustrated in Chart 1 and the invocation of an *ortho*-oxygenation process. Thus from Chart 1 the key intermediate prephenic acid can lead to either the nonoxygenated aromatic series (e.g., cinnamic acid) or the *para*-hydroxy series exemplified by *p*-coumaric acid; *ortho*-hydroxylation of the latter leads to the familiar 3,4-dihydroxy series (e.g., caffeic acid) and further to the 3,4,5-trihydroxy derivatives. It should be mentioned that the 3,4-di- and 3,4,5-trioxygenated benzoic acids can also be derived from 5-dehydroshikimic acid directly. An aspect of oxygenation which is particularly significant in coumarin biosynthesis is the introduction of an oxygen function at the *ortho* position relative to the side chain. Such free phenols are not common but there can be some ambiguity in biogenetic classification as a result of this process. Salicylic acid, for example, with its single *ortho*-hydroxyl function is difficult to rationalize biogenetically except as an oxygenated shikimate-derived benzoic acid. It has been a general policy in dealing with the phenols to place those in the acetogenic classes which most obviously belong there, and to put the remainder into the shikimate classes. Thus, some unusual oxygenation patterns will be found which cannot necessarily be rationalized as discussed above. In the absence of biosynthetic evidence, such ambiguities have to be accepted, although from time to time biogenetic models are presented which shed some light on the origins of such patterns—e.g., the rearrangement mechanism leading to gentisic acid:



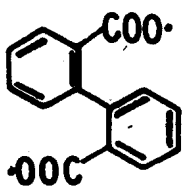
SHIKIMATE CLASSES

Ar* =  or or or

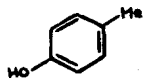
* See page 5

01	C_6C_1	Ar—C
02	C_6C_2	Ar—C—C
03	C_6C_3	Ar—C—C—C
04	C_6C_n	Ar—[C]$_n$ n>4
05	Coumarins	
06	Lignans	
07	Terphenyls	
1902	Prearomatics	e.g. 

01	C_6C_1	Ar–C

0101	$Ar-CH_3$
0102	$Ar-CH_2O\cdot$
0103	Ar–CHO
0104	$Ar-COO\cdot$
0105	
0199	MISC.

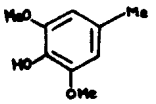
CRESOL, p- C_7H_8O 108 OI 36°C



Chem. Abstr., 1956 (50) 9693.

0101-001

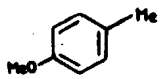
PYROGALLOL-1,3-DI-ME ETHER, -5-Me- $C_9H_{12}O_3$ 168 OI 41°C



Ber., 1936 (69) 1870.

0101-006

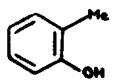
CRESOL ME ETHER, p- $C_8H_{10}O$ 122 OI Oil



Helv., 1964, 111.

0101-002

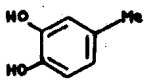
CRESOL, ortho- C_7H_8O 108 OI Oil



Phytochem., 1968, 278.

0101-007

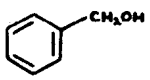
HOMOCATECHOL $C_7H_8O_2$ 124 OI 65°C



Chem. Zentr., 1931 (II) 586.

0101-003

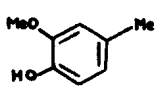
BENZYL ALCOHOL C_7H_8O 108 OI Oil



Arch. Pharm., 1932 (270) 249.

0102-001

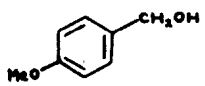
CREOSOL $C_8H_{10}O_2$ 138 OI Oil



Phytochemistry, 1968, 278.

0101-004

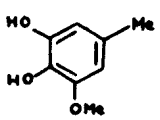
ANISYL ALCOHOL $C_8H_{10}O_2$ 138 OI 25°C



Chem. Abstr., 1951, 3124.

0102-002

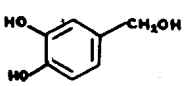
PYROGALLOL-1-ME ETHER, -5-Me- $C_8H_{10}O_3$ 154 OI 68°C



Ber., 1936, 1870.

0101-005

BENZYL ALCOHOL, 3,4-diOH- $C_7H_8O_3$ 140 OI



Phytochem., 1968, 119.

0102-003