Medicinal Natural Products

A Biosynthetic Approach
3rd Edition

Paul M Dewick

formerly University of Nottingham, UK

WILEY

A John Wiley and Sons, Ltd., Publication

Medicinal Natural Products

Medicinal Natural Products

A Biosynthetic Approach
3rd Edition

Paul M Dewick

formerly University of Nottingham, UK

WILEY

A John Wiley and Sons, Ltd., Publication

This edition first published 2009

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Dewick, Paul M.
Medicinal natural products : a biosynthetic approach / Paul M. Dewick. – 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-74168-9 (cloth) – ISBN 978-0-470-74167-2 (pbk. : alk. paper)
1. Natural products. 2. Biosynthesis. 3. Pharmacognosy. I. Title.
RS160.D48 2008
615′.3 – dc22

2008044502

A catalogue record for this book is available from the British Library.

ISBN: 978-0-470-74168-9 (HB) 978-0-470-74167-2 (PB)

Typeset in 9pt/11pt Times Roman by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire

Cover photograph
English yew (*Taxus baccata*). Photograph: Paul Dewick

for Jane

CONTENTS

1

ABOUT THIS BOOK, AND HOW TO USE IT

THE SUBJECT

This book has been written primarily for pharmacy students to provide a modern text to complement lecture courses dealing with pharmacognosy and the use of natural products in medicine. Nevertheless, it should be of value in other courses where the study of natural products is included, although the examples chosen are predominantly those possessing pharmacological activity.

For centuries, drugs were entirely of natural origin and composed of herbs, animal products, and inorganic materials. Early remedies may have combined these ingredients with witchcraft, mysticism, astrology, or religion, but it is certain that those treatments that were effective were subsequently recorded and documented, thus leading to the early Herbals. The science of pharmacognosy – the knowledge of drugs – grew from these records to provide a disciplined, scientific description of natural materials used in medicine. Herbs formed the bulk of these remedies. As chemical techniques improved, the active constituents were isolated from plants, were structurally characterized, and, in due course, many were synthesized in the laboratory. Sometimes, more active, better-tolerated drugs were produced by chemical modifications (semi-synthesis), or by total synthesis of analogues of the active principles.

Gradually, synthetic compounds superseded many of the old plant drugs, though certain plant-derived agents were never surpassed and remain as valued medicines to this day. Natural drugs derived from microorganisms have a much shorter history, and their major impact on medicine goes back only about 60 years to the introduction of the antibiotic penicillin. Microbially produced antibiotics now account for a very high proportion of the drugs commonly prescribed. There is currently a renewed interest in pharmacologically active natural products, be they from plants, microorganisms, or animals, terrestrial or marine, in the continued search for new drugs, particularly for disease states where our present range of drugs is less effective than we would wish. This is being reflected in a growing number of natural products or natural-product-inspired drugs entering medicine. Herbal remedies are also enjoying a revival as many sufferers turn away from modern drugs and embrace 'complementary medicine'.

THE AIM

Many university pharmacy courses include a pharmacognosy component covering a study of plant-derived drugs; traditionally, this area of natural products has been taught separately from the microbially derived antibiotics, or the animal-related steroidal and prostanoid drugs. Such topics have usually formed part of a pharmaceutical chemistry course. The traditional boundaries may still remain, despite a general change in pharmacognosy teaching from a descriptive study to a phytochemical-based approach, a trend towards integrating pharmacognosy within pharmaceutical chemistry, and the general adoption of modular course structures. A chemistry-based teaching programme encompassing all types of natural products of medicinal importance, semi-synthetic derivatives, and synthetic analogues based on natural product templates is a logical development. This book provides a suitable text to complement such a programme, and attempts to break down the artificial divisions.

Medicinal Natural Products: A Biosynthetic Approach. 3rd Edition Paul Dewick

THE APPROACH

This book provides a groundwork in natural product chemistry/phytochemistry by considering biosynthesis – the metabolic sequences leading to various selected classes of natural products. This allows application of fundamental chemical principles and displays the relationships between the diverse structures encountered in nature, thus providing a rationale for natural products and replacing a descriptive approach with one based more on deductive reasoning. It also helps to transform complicated structures into a comprehensible combination of simpler fragments; natural product structures can be quite complex. Subdivision of the topics is predominantly via biosynthesis, not by class or activity, and this provides a logical sequence of structural types and avoids a catalogue effect. There is extensive use of chemical schemes and mechanism, with detailed mechanistic explanations being annotated to the schemes, as well as outline discussions in the text. Lots of cross-referencing is included to emphasize links and similarities; it is not necessary to follow these to understand the current material, but they are used to stress that the concept has been met before, or that other uses will be met in due course. As important classes of compounds or drugs are reached, more detailed information is then provided in the form of short separate monographs in boxes, which can be studied or omitted as required, in the latter case allowing the main theme to continue. The monograph information covers sources, production methods, principal components, drug use, mode of action, semi-synthetic derivatives, synthetic analogues, etc., as appropriate. Those materials currently employed as drugs, or being tested clinically, are emphasized in the monographs by the use of bold type.

THE TOPICS

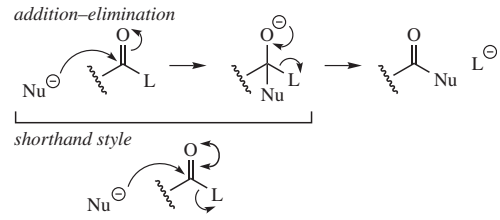
A preliminary chapter is used to outline the main building blocks, the basic construction mechanisms employed in the biosynthesis of natural products, and how metabolic pathways are deduced. Most of the fundamental principles should be familiar and will have been met previously in courses dealing with the basics of organic chemistry and biochemistry. These principles are then seen in action as representative natural product structures are described in the following chapters. The topics selected are subdivided initially into areas of metabolism fed by the acetate, shikimate, mevalonate, and methylerythritol phosphate pathways. The remaining chapters then cover alkaloids, peptides and proteins, and carbohydrates. Not all classes of natural products can be covered, and the book is intended as an introductory text, not a comprehensive reference work.

The book tries to include a high proportion of those natural products currently used in medicine, the major drugs that are derived from natural materials by semi-synthesis, and those drugs which are structural analogues. Some of the compounds mentioned may have a significant biological activity which is of interest, but not medicinally useful. The book is also designed to be forward looking and gives information on possible leads to new drugs and materials in clinical trials.

THE FIGURES

A cursory glance through the book will show that a considerable portion of the content is in the form of chemical structures and schemes. The schemes and figures are used to provide maximum information as concisely as possible. The following guidelines should be appreciated:

- A figure may present a composite scheme derived from studies in more than one organism.
- Comments in italics provide an explanation in chemical terms for the biochemical reaction; detailed enzymic mechanisms are not usually considered.
- Schemes in separate frames show a mechanism for part of the sequence, the derivation of a substrate, or perhaps structurally related systems.
- Although enzymic reactions may be reversible, single rather than reversible arrows are used, unless the transformation is one that may be implicated in both directions, e.g. amino acid/keto acid transaminations.
- E1, E2, etc., refer to enzymes catalysing the transformation, when known. Where no enzyme is indicated, the transformation may well have been determined by other methodology, e.g. isotope tracer studies. Speculative conversions may be included, but are clearly indicated.
- Enzyme names shown are the commonly accepted names; in general, only one name is given, even though alternative names may also be in current use.
- Proteins identified via the corresponding gene are often assigned a code name/number by researchers, and no systematic name has been proposed. This means that proteins carrying out the same transformation in different organisms may be assigned different codes.
- Double-headed curly arrows are used to represent an addition–elimination mechanism as follows:



FURTHER READING

A selection of articles suitable for supplementary reading is provided at the end of each chapter. In general, these are not chosen from the primary literature, but are recent review articles covering broader aspects of the topic. They are also located in easily accessible journals rather than books, and have been chosen as the most student friendly. In certain cases, the most recent reviews available may be somewhat less up to date than the information covered in this book. All of the selected articles contain information considered appropriate to this book, e.g. reviews on 'synthesis' may contain sections on structural aspects, biosynthesis, or pharmacology.

WHAT TO STUDY

Coverage is fairly extensive to allow maximum flexibility for courses in different institutions, and not all of the material will be required for any one course. However, because of the many subdivisions and the highlighted keywords, it should be relatively easy to find and select the material appropriate for a particular course. On the other hand, the detail given in monographs is purposely limited to ensure students are provided with enough factual information, but are not faced with the need to assess whether or not the material is relevant. Even so, these monographs will undoubtedly contain data which exceed the scope of any individual course. It is thus necessary to apply selectivity, and portions of the book will be surplus to immediate requirements. The book is designed to be user friendly, suitable for modular courses and student-centred learning exercises, and a starting point for later project and dissertation work. The information presented is as up to date as possible; undoubtedly, new research will be published that modifies or even contradicts some of the statements made. The reader is asked always to be critical and to maintain a degree of flexibility when reading the scientific literature, and to appreciate that science is always changing.

WHAT TO LEARN

The primary aim of the book is not to rely just on factual information, but to impart an understanding of natural product structures and the way they are put together by living organisms. Rationalization based on mechanistic reasoning is paramount. The sequences themselves are not important, whilst the names of chemicals and the enzymes involved in the pathways are even less relevant and included only for information; it is the mechanistic explanations that are the essence. Students should concentrate on understanding the broad features of the sequences and absorb sufficient information to be able to predict how and why intermediates might be elaborated and transformed. The mechanistic explanations appended to the schemes should reinforce this approach. Anyone who commits to memory a sequence of reactions for examination purposes has missed the point. There is no alternative to memory for some of the material covered in the monographs, if it is required; wherever possible, information should be reduced to a concept that can be deduced, rather than remembered. The approach used here should help students to develop such deductive skills.

NOMENCLATURE

Natural product structures are usually quite complex, some exceedingly so, and fully systematic nomenclature becomes impracticable. Names are thus typically based on so-called trivial nomenclature, in which the discoverer of the natural product exerts their right to name the compound. The organism in which the compound is found is frequently chosen to supply the root name, e.g. hyoscyamine from *Hyoscyamus*, atropine from *Atropa*, or penicillin from *Penicillium*. Name suffixes might be -in to indicate 'a constituent of', -oside to show the compound is a sugar derivative, -genin for the aglycone released by hydrolysis of the sugar derivative, -toxin for a poisonous constituent, or may reflect chemical functionality, such as -one or -ol. Traditionally, -ine is always used for alkaloids (am*ine*s).

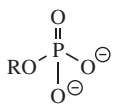
Structurally related compounds are then named as derivatives of the original, using standard prefixes, such as hydroxy-, methoxy-, methyl-, dihydro-, homo-, etc. for added substituents, or deoxy-, demethyl-, demethoxy-, dehydro-, nor-, etc. for removed substituents. Homo- is used to indicate one carbon more, whereas nor- means one carbon less. The position of this change is then indicated by systematic numbering of the carbon chains or rings. Some groups of compounds, such as steroids, fatty acids, and prostaglandins, are named semi-systematically from an accepted root name for the complex hydrocarbon skeleton. In this book, almost all structures depicted in the figures carry a name; this is primarily to help identification, and, for the student, structural features should be regarded as more pertinent than the names used.

It will soon become apparent that drug names chosen by pharmaceutical manufacturers are quite random, and in most cases have no particular relationship to the chemical structure. However, some common stems are employed to indicate relationship to a group of therapeutically active drugs. Examples are -cillin for antibiotics of the penicillin group, cef- for antibiotics of the cephalosporin group, -mycin for antibiotics produced

by *Streptomyces*, -caine for local anaesthetics, -stat for enzyme inhibitors, -vastatin for HMGCoA reductase inhibitors, prost for prostaglandins, and gest for progestogens. We are also currently still in a transitional period during which many established drug names are being changed to recommended international non-proprietary names (rINNs); both names are included here, with the rINN preceding the older name.

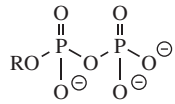
CONVENTIONS REGARDING ACIDS, BASES, AND IONS

In many structures, the abbreviation O**P** is used to represent the phosphate group and O**PP** the diphosphate (or pyrophosphate) group:



a phosphate

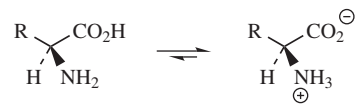
RO**P**



a diphosphate (pyrophosphate)

RO**PP**

At physiological pH values, these groups will be ionized as shown, but in schemes where structures are given in full, the non-ionized acids are usually depicted. This is done primarily to simplify structures, to eliminate the need for counter-ions, and to avoid mechanistic confusion. Likewise, amino acids are shown in non-ionized form, although they will typically exist as zwitterions:



L-amino acid zwitterion

Ionized and non-ionized forms of many compounds are regarded as synonymous in the text; thus, acetate/acetic acid, shikimate/shikimic acid, and mevalonate/mevalonic acid may be used according to the author's whim and context and have no especial relevance.

SOME COMMON ABBREVIATIONS

5-HT	5-hydroxytryptamine
ACP	acyl carrier protein
ADP	adenosine diphosphate
Ara	arabinose
ATP	adenosine triphosphate
B:	general base
CoA	coenzyme A as part of a thioester, e.g. acetyl-CoA ($CH_3COSCoA$)
Dig	digitoxose
DMAPP	dimethylallyl diphosphate (dimethylallyl pyrophosphate)
DXP	1-deoxyxylulose 5-phosphate
Enz	enzyme (usually shown as thiol: EnzSH)
FAD	flavin adenine dinucleotide
$FADH_2$	flavin adenine dinucleotide (reduced)
FAS	fatty acid synthase
FH_4	tetrahydrofolic acid
FMN	flavin mononucleotide
$FMNH_2$	flavin mononucleotide (reduced)
FPP	farnesyl diphosphate (farnesyl pyrophosphate)
Fru	fructose
GABA	γ-aminobutyric acid
Gal	galactose
GFPP	geranylfarnesyl diphosphate (geranylfarnesyl pyrophosphate)
GGPP	geranylgeranyl diphosphate (geranylgeranyl pyrophosphate)
Glc	glucose
GPP	geranyl diphosphate (geranyl pyrophosphate)
HA	general acid
HSCoA	coenzyme A
IPP	isopentenyl diphosphate (isopentenyl pyrophosphate)
LT	leukotriene
Mann	mannose
MEP	methylerythritol phosphate
MVA	mevalonic acid
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NRPS	non-ribosomal peptide synthase
O	oxidation – in schemes
P	phosphate – in text
P	phosphate – in structures
PCP	peptidyl carrier protein
PEP	phosphoenolpyruvate
PG	prostaglandin
PKS	polyketide synthase
PLP	pyridoxal 5′-phosphate
PP	diphosphate (pyrophosphate) – in text
PP	diphosphate (pyrophosphate) – in structures
Rha	rhamnose
Rib	ribose

SAM	*S*-adenosyl methionine
TPP	thiamine diphosphate (thiamine pyrophosphate)
TX	thromboxane
UDP	uridine diphosphate
UDPGlc	uridine diphosphoglucose
UTP	uridine triphosphate
W–M	Wagner–Meerwein (rearrangement)
Xyl	xylose
Δ	heat
$h\nu$	electromagnetic radiation; usually UV or visible

FURTHER READING

Pharmacognosy, Phytochemistry, Natural Drugs

Books

Bruneton J (1999) *Pharmacognosy, Phytochemistry and Medicinal Plants*. Lavoisier, Paris.

Evans WC (2001) *Trease & Evans' Pharmacognosy*. Saunders, London.

Heinrich M, Barnes J, Gibbons S and Williamson EM (2004) *Fundamentals of Pharmacognosy and Phytotherapy*. Churchill Livingstone, London.

Samuelsson G (2004) *Drugs of Natural Origin. A Textbook of Pharmacognosy*. Swedish Pharmaceutical Press, Stockholm.

Reviews

Baker DD, Chu M, Oza U and Rajgarhia V (2007) The value of natural products to future pharmaceutical discovery. *Nat Prod Rep* **24**, 1225–1244.

Bode HB and Müller R (2005) The impact of bacterial genomics on natural product research. *Angew Chem Int Ed* **44**, 6828–6846.

Butler MS (2004) The role of natural product chemistry in drug discovery. *J Nat Prod* **67**, 2141–2153; errata (2006) **69**, 172.

Butler MS (2008) Natural products to drugs: natural product-derived compounds in clinical trials. *Nat Prod Rep* **25**, 475–516.

Chin Y-W, Balunas MJ, Chai HB and Kinghorn AD (2006) Drug discovery from natural sources. *AAPS J* **8**, 239–253.

Espín JC, García-Conesa MT and Tomás-Barberán FA (2007) Nutraceuticals: facts and fiction. *Phytochemistry* **68**, 2986–3008.

Kennedy J (2008) Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat Prod Rep* **25**, 25–34.

McChesney JD, Venkataraman SK and Henri JT (2007) Plant natural products: back to the future or into extinction? *Phytochemistry* **68**, 2015–2022.

Misiek M and Hoffmeister D (2007) Fungal genetics, genomics, and secondary metabolites in pharmaceutical sciences. *Planta Med* **73**, 103–115.

Newman DJ (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery? *J Med Chem* **51**, 2589–2599.

Newman DJ and Cragg GM (2004) Marine natural products and related compounds in clinical and advanced preclinical trials. *J Nat Prod* **67**, 1216–1238.

Newman DJ and Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **70**, 461–477.

Phillipson JD (2007) Phytochemistry and pharmacognosy. *Phytochemistry* **68**, 2960–2972.

Tietze LF, Bell HP and Chandrasekhar S (2003) Natural product hybrids as new leads for drug discovery. *Angew Chem Int Ed* **42**, 3996–4028.

2

SECONDARY METABOLISM: THE BUILDING BLOCKS AND CONSTRUCTION MECHANISMS

PRIMARY AND SECONDARY METABOLISM

All organisms need to transform and interconvert a vast number of organic compounds to enable them to live, grow, and reproduce. They need to provide themselves with energy in the form of ATP, and a supply of building blocks to construct their own tissues. An integrated network of enzyme-mediated and carefully regulated chemical reactions is used for this purpose, collectively referred to as **intermediary metabolism**, and the pathways involved are termed **metabolic pathways**. Some of the crucially important molecules of life are carbohydrates, proteins, fats, and nucleic acids. Apart from fats, these tend to be polymeric materials. Carbohydrates are composed of sugar units, whilst proteins are made up from amino acids, and nucleic acids are based on nucleotides. Organisms vary widely in their capacity to synthesize and transform chemicals. For instance, plants are very efficient at synthesizing organic compounds via photosynthesis from inorganic materials found in the environment, whilst other organisms, such as animals and microorganisms, rely on obtaining their raw materials in their diet, e.g. by consuming plants. Thus, many of the metabolic pathways are concerned with degrading materials taken in as food, whilst others are then required to synthesize specialized molecules from the basic compounds so obtained.

Despite the extremely varied characteristics of living organisms, the pathways for generally modifying and synthesizing carbohydrates, proteins, fats, and nucleic acids are found to be essentially the same in all organisms, apart from minor variations. These processes demonstrate the fundamental unity of all living matter, and are collectively described as **primary metabolism**, with the compounds involved in the pathways being termed **primary metabolites**. Thus, degradation of carbohydrates and sugars generally proceeds via the well-characterized pathways known as glycolysis and the Krebs/citric acid/tricarboxylic acid cycle, which release energy from the organic compounds by oxidative reactions. Oxidation of fatty acids from fats by the sequence called β-oxidation also provides energy. Aerobic organisms are able to optimize these processes by adding on a further process, namely oxidative phosphorylation. This improves the efficiency of oxidation by incorporating a more general process applicable to the oxidation of a wide variety of substrates rather than having to provide specific processes for each individual substrate. Proteins taken in via the diet provide amino acids, but the proportions of each will almost certainly vary from the organism's requirements. Metabolic pathways are thus available to interconvert amino acids, or degrade those not required and thus provide a further source of energy. Most organisms can synthesize only a proportion of the amino acids they actually require for protein synthesis. Those structures not synthesized, so-called essential amino acids, must be obtained from external sources.

In contrast to these primary metabolic pathways, which synthesize, degrade, and generally interconvert compounds commonly encountered in all organisms, there also exists an area of metabolism concerned with compounds which have a much more limited distribution in nature. Such compounds, called **secondary metabolites**,

Medicinal Natural Products: A Biosynthetic Approach. 3rd Edition Paul Dewick

are found in only specific organisms, or groups of organisms, and are an expression of the individuality of species. Secondary metabolites are not necessarily produced under all conditions, and in the vast majority of cases the function of these compounds and their benefit to the organism are not yet known. Some are undoubtedly produced for easily appreciated reasons, e.g. as toxic materials providing defence against predators, as volatile attractants towards the same or other species, or as colouring agents to attract or warn other species, but it is logical to assume that all do play some vital role for the well-being of the producer. It is this area of **secondary metabolism** which provides most of the pharmacologically active natural products. It is thus fairly obvious that the human diet could be both unpalatable and remarkably dangerous if all plants, animals, and fungi produced the same range of compounds.

The above generalizations distinguishing primary and secondary metabolites unfortunately leave a 'grey area' at the boundary, so that some groups of natural products could be assigned to either division. Fatty acids and sugars provide good examples, in that most are best described as primary metabolites, whilst some representatives are extremely rare and found only in a handful of species. Likewise, steroid biosynthesis produces a range of widely distributed fundamental structures, yet some steroids, many of them with pronounced pharmacological activity, are restricted to certain organisms. Hopefully, the blurring of the boundaries will not cause confusion; the subdivision into primary metabolism (≡ biochemistry) or secondary metabolism (≡ natural products chemistry) is merely a convenience and there is considerable overlap.

THE BUILDING BLOCKS

The building blocks for secondary metabolites are derived from primary metabolism as indicated in Figure 2.1. This scheme outlines how metabolites from the fundamental processes of photosynthesis, glycolysis, and the Krebs cycle are tapped off from energy-generating processes to provide biosynthetic intermediates. The number of building blocks needed is surprisingly few, and as with any child's construction set, a vast array of objects can be built up from a limited number of basic building blocks. By far the most important building blocks employed in the biosynthesis of secondary metabolites are derived from the intermediates acetyl coenzyme A (acetyl-CoA), shikimic acid, mevalonic acid, and methylerythritol phosphate. These are utilized respectively in the **acetate, shikimate, mevalonate**, and **methylerythritol phosphate** pathways, which form the basis of succeeding chapters. **Acetyl-CoA** is formed by oxidative decarboxylation of the glycolytic pathway product pyruvic acid. It is also produced by the β-oxidation of fatty acids, effectively reversing the process by which fatty acids are themselves synthesized from acetyl-CoA. Important secondary metabolites formed from the acetate pathway include phenols, prostaglandins, and macrolide antibiotics, together with various fatty acids and derivatives at the primary–secondary metabolism interface. **Shikimic acid** is produced from a combination of phosphoenolpyruvate, a glycolytic pathway intermediate, and erythrose 4-phosphate from the pentose phosphate pathway. The reactions of the pentose phosphate cycle may be employed for the degradation of glucose, but they also feature in the synthesis of sugars by photosynthesis. The shikimate pathway leads to a variety of phenols, cinnamic acid derivatives, lignans, and alkaloids. **Mevalonic acid** is itself formed from three molecules of acetyl-CoA, but the mevalonate pathway channels acetate into a different series of compounds than does the acetate pathway. **Methylerythritol phosphate** arises from a combination of two glycolytic pathway intermediates, namely pyruvic acid and glyceraldehyde 3-phosphate by way of deoxyxylulose phosphate. The mevalonate and methylerythritol phosphate pathways are together responsible for the biosynthesis of a vast array of terpenoid and steroid metabolites.

In addition to acetyl-CoA, shikimic acid, mevalonic acid, and methylerythritol phosphate, other building blocks based on amino acids are frequently employed in natural product synthesis. Peptides, proteins, alkaloids, and many antibiotics are derived from amino acids, and the origins of some of the more important amino acid components of these are briefly indicated in Figure 2.1. Intermediates from the glycolytic pathway and the Krebs cycle are used in constructing many of them, but the aromatic amino acids **phenylalanine, tyrosine**, and **tryptophan** are themselves products from the shikimate pathway. **Ornithine**, an amino acid not found in proteins, and its homologue **lysine**, are important alkaloid precursors and have their origins in Krebs cycle intermediates.

Of special significance is the appreciation that secondary metabolites can be synthesized by combining several building blocks of the same type, or by using a mixture of different building blocks. This expands structural diversity and, consequently, makes subdivisions based entirely on biosynthetic pathways rather more difficult. A typical natural product might be produced by combining elements from the acetate, shikimate, and methylerythritol phosphate pathways, for example. Many secondary metabolites also contain one or more sugar units in their structure, either simple primary metabolites,

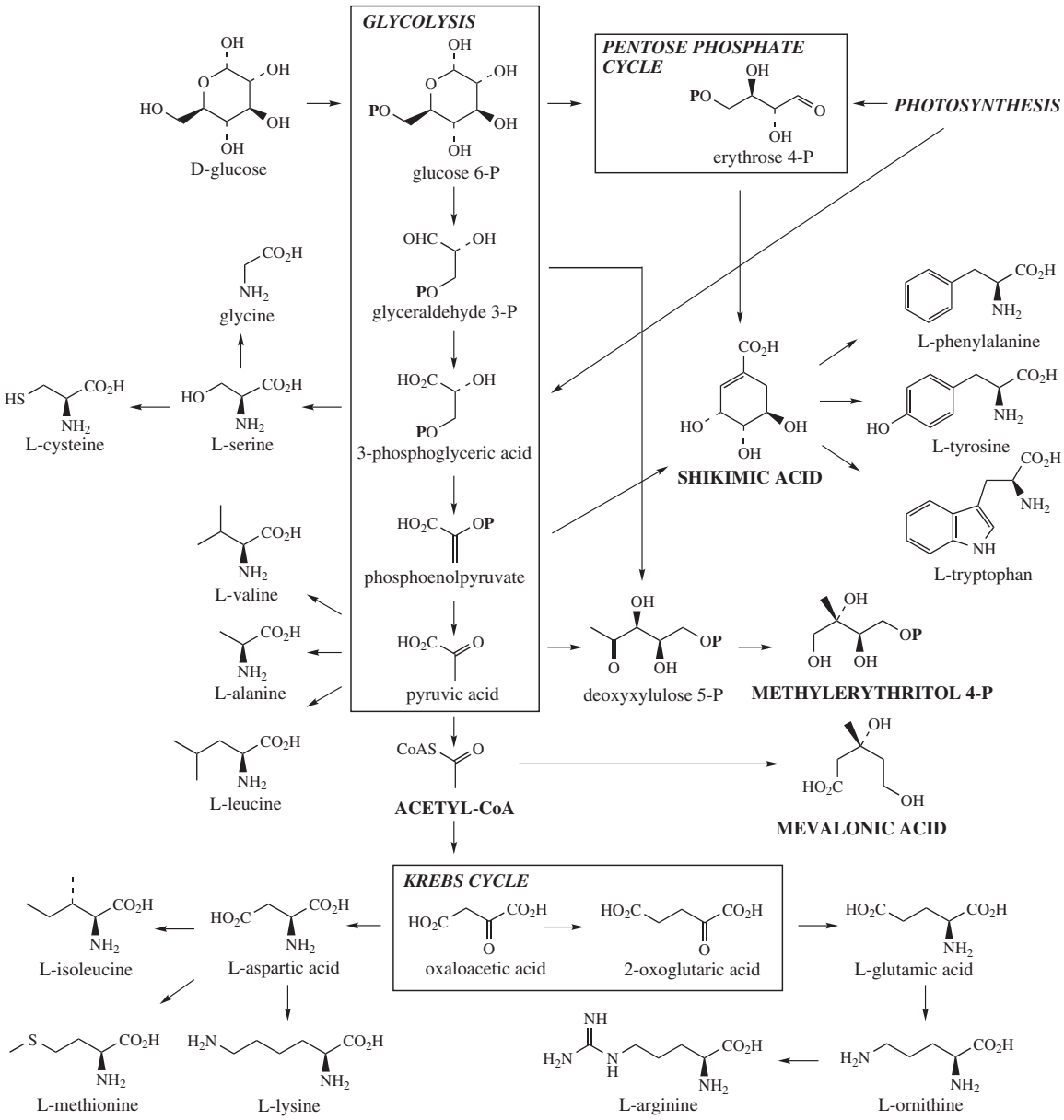


Figure 2.1

such as glucose or ribose, or alternatively substantially modified and unusual sugars. To appreciate how a natural product is elaborated, it is of value to be able to dissect its structure into the basic building blocks from which it is made up and to use fundamental chemical mechanisms to propose how these are joined together. With a little experience and practice, this becomes a relatively simple process and it allows the molecule to be rationalized, thus exposing logical relationships between apparently quite different structures. In this way, similarities become much more meaningful than differences, and an understanding of biosynthetic pathways allows rational connecting links to be established. This forms the basic approach in this book.

Relatively few building blocks are routinely employed, and the following list, though not comprehensive, includes those most frequently encountered in producing the carbon and nitrogen skeleton of a natural product. As we shall see, oxygen atoms can be introduced and removed by a variety of processes, and so are not considered in the initial analysis, except as a pointer to an acetate (see page 101) or shikimate (see page 140) origin. The structural features of these building blocks are shown in Figure 2.2.

- **C_1:** The simplest of the building blocks is composed of a single carbon atom, usually in the form of a methyl group, and most frequently it is attached to oxygen or nitrogen, but occasionally to carbon or sulfur. It is derived from the *S*-methyl of L-**methionine**. The methylenedioxy group (–OCH_2O–) is also an example of a C_1 unit.
- **C_2:** A two-carbon unit may be supplied by **acetyl-CoA**. This could be a simple acetyl group, as in an ester, but more frequently it forms part of a long alkyl chain (as in a fatty acid) or may be part of an aromatic system (e.g. phenols). Of particular relevance is that, in the latter examples, acetyl-CoA is first converted into the more reactive **malonyl-CoA** before its incorporation.
- **C_5:** The branched-chain C_5 'isoprene' unit is a feature of compounds formed from **mevalonate** or

The building blocks

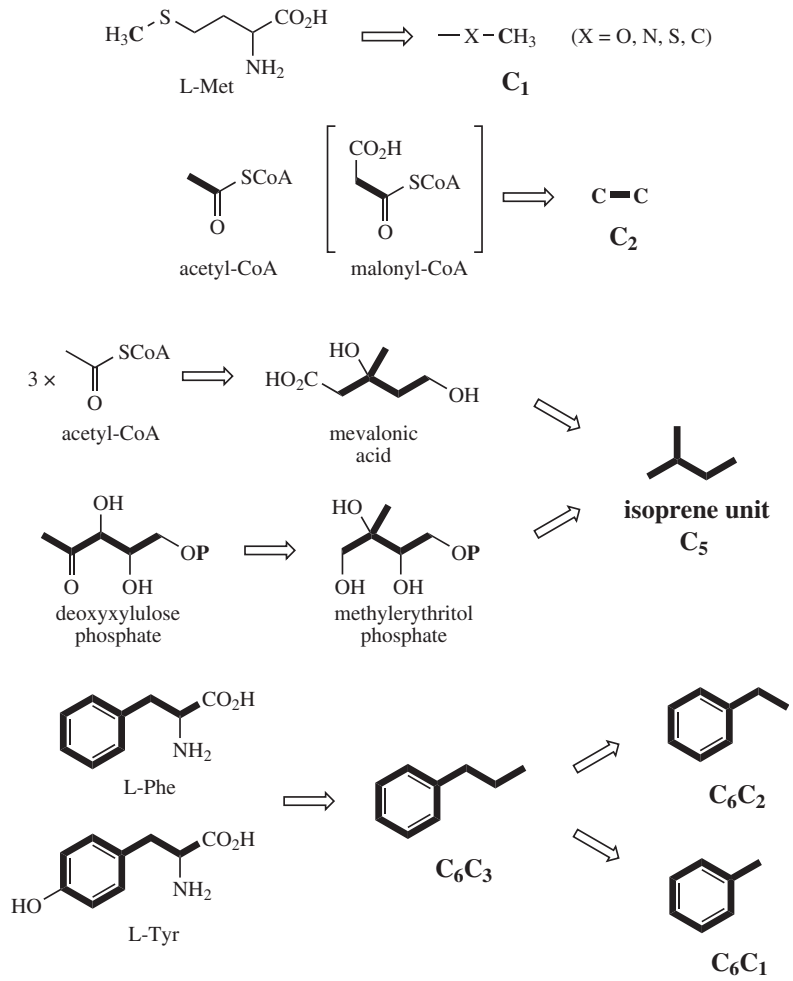


Figure 2.2

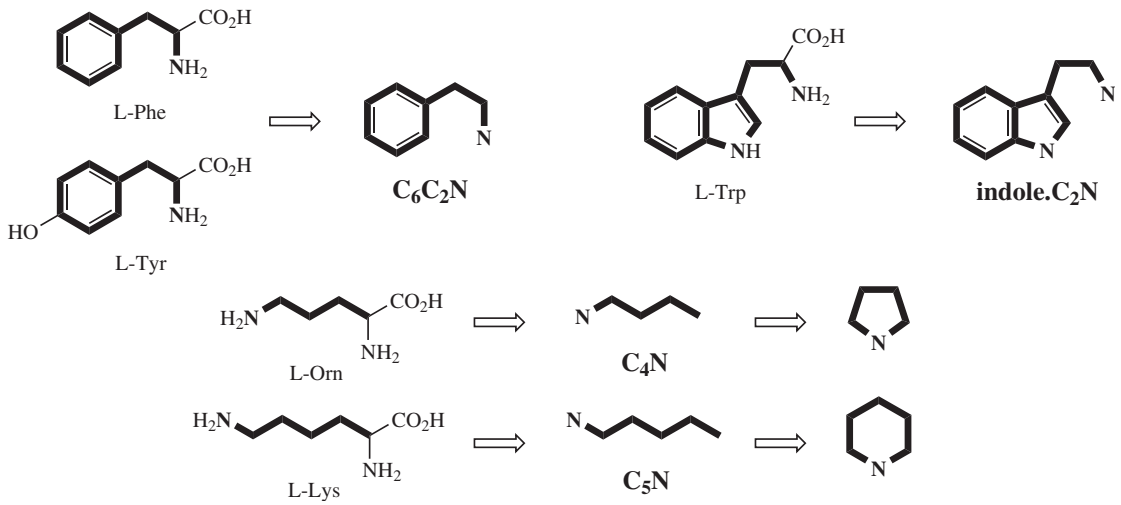


Figure 2.2 *(Continued)*

methylerythritol phosphate. Mevalonate itself is the product from three acetyl-CoA molecules, but only five of mevalonate's six carbon atoms are used, the carboxyl group being lost. The alternative precursor, methylerythritol phosphate, is formed from a straight-chain sugar derivative, deoxyxylulose phosphate, which undergoes a skeletal rearrangement to form the branched-chain isoprene unit.

- **C_6C_3:** This refers to a phenylpropyl unit and is obtained from the carbon skeleton of either L-**phenylalanine** or L-**tyrosine**, two of the shikimate-derived aromatic amino acids. This, of course, requires loss of the amino group. The C_3 side-chain may be saturated or unsaturated, and may be oxygenated. Sometimes the side-chain is cleaved, removing one or two carbon atoms. Thus, $\mathbf{C_6C_2}$ and $\mathbf{C_6C_1}$ units represent modified shortened forms of the C_6C_3 system.
- **C_6C_2N:** Again, this building block is formed from either L-**phenylalanine** or L-**tyrosine**, with L-tyrosine being by far the more common precursor. In the elaboration of this unit, the carboxyl carbon of the amino acid is removed.
- **indole.C_2N:** The third of the aromatic amino acids is L-**tryptophan**. This indole-containing system can undergo decarboxylation in a similar way to L-phenylalanine and L-tyrosine, so providing the remainder of the skeleton as an indole.C_2N unit.
- **C_4N:** The C_4N unit is usually found as a heterocyclic pyrrolidine system and is produced from the non-protein amino acid L-**ornithine**. In marked contrast to the C_6C_2N and indole.C_2N units described above, ornithine supplies not its α-amino nitrogen, but the δ-amino nitrogen. The carboxylic acid function and the α-amino nitrogen are both lost.
- **C_5N:** This is produced in exactly the same way as the C_4N unit, but using L-**lysine** as precursor. The ε-amino nitrogen is retained, and the unit is commonly found as a piperidine ring system.

These eight building blocks will form the basis of many of the natural product structures discussed in the following chapters. Simple examples of how compounds can be visualized as a combination of building blocks are shown in Figure 2.3. At this stage, it is inappropriate to justify why a particular combination of units is used, but this aspect should become clear as the pathways are described. A word of warning is also necessary. Some natural products have been produced by processes in which a fundamental rearrangement of the carbon skeleton has occurred. This is especially common with structures derived from isoprene units, and it obviously disguises some of the original building blocks from immediate recognition. The same is true if one or more carbon atoms are removed by oxidation reactions.

THE CONSTRUCTION MECHANISMS

Natural product molecules are biosynthesized by a sequence of reactions which, with very few exceptions, are catalysed by enzymes. Enzymes are protein molecules which facilitate chemical modification of substrates by virtue of their specific binding properties conferred by

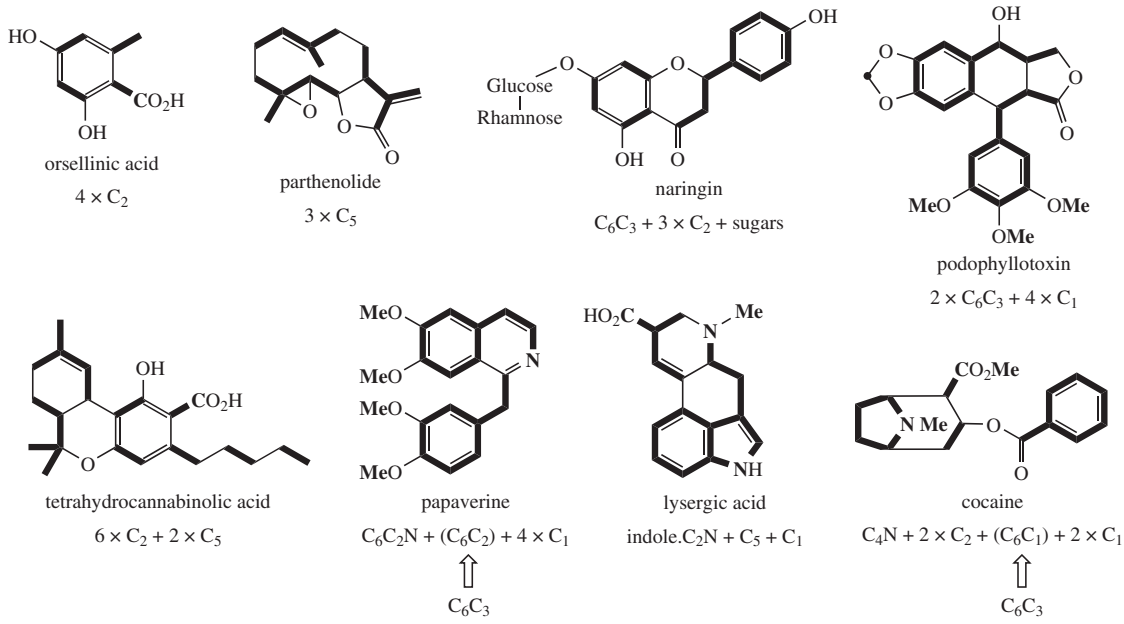


Figure 2.3

the particular combination of functional groups in the constituent amino acids. In many cases, a suitable cofactor, e.g. NAD^+, PLP, HSCoA (see below), as well as the substrate, may also be bound to participate in the transformation. Although enzymes catalyse some fairly elaborate and sometimes unexpected changes, it is generally possible to account for the reactions using established chemical principles and mechanisms. As we explore the pathways to a wide variety of natural products, the reactions will generally be discussed in terms of chemical analogies. Enzymes have the power to effect these transformations more efficiently and more rapidly than the chemical analogy, and also under very much milder conditions. Where relevant, they also carry out reactions in a stereospecific manner. Some of the important reactions frequently encountered are now described.

Alkylation Reactions: Nucleophilic Substitution

The C_1 methyl building unit is supplied from L-methionine and is introduced by a nucleophilic substitution reaction. In nature, the leaving group is enhanced by converting L-methionine into ***S*-adenosylmethionine** (**SAM**, AdoMet) [Figure 2.4(a)]. This gives a positively charged sulfur and facilitates the S_N2-type nucleophilic substitution mechanism [Figure 2.4(b)]. Thus, *O*-methyl and *N*-methyl linkages may be obtained using hydroxyl and amino functions as nucleophiles. Methionine is subsequently regenerated by the methylation of homocysteine, using N^5-methyl-tetrahydrofolate (see page 144) as methyl donor. The generation of *C*-methyl linkages requires the participation of nucleophilic carbon. Positions *ortho* or *para* to a phenol group, or positions adjacent to one or more carbonyl groups, are thus candidates for *C*-methylation [Figure 2.4(c)].

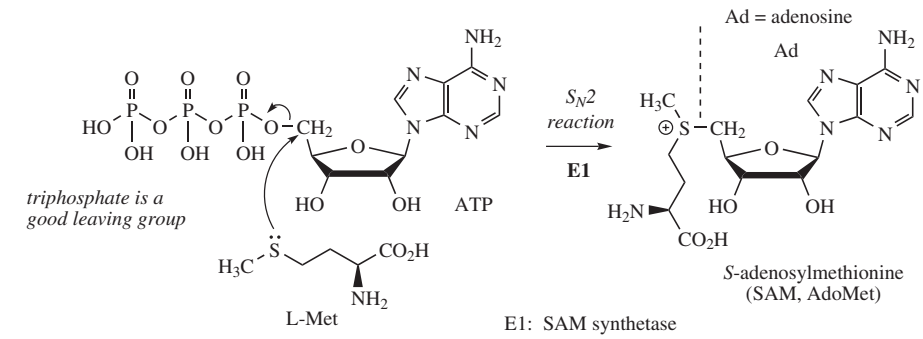
A C_5 isoprene unit in the form of **dimethylallyl diphosphate** (**DMAPP**) may also act as an alkylating agent, since diphosphate is a good leaving group [Figure 2.4(d)]. Although a similar S_N2 displacement may be proposed, the available evidence points to an S_N1 process. DMAPP first ionizes to the resonance-stabilized allylic carbocation, and the nucleophile is able to attack either of the cationic centres. In the majority of cases, the nucleophile attacks the same carbon that loses the diphosphate. *C*-Alkylation at activated positions using DMAPP is analogous to the *C*-methylation process above, though by an S_N1 mechanism.

Alkylation Reactions: Electrophilic Addition

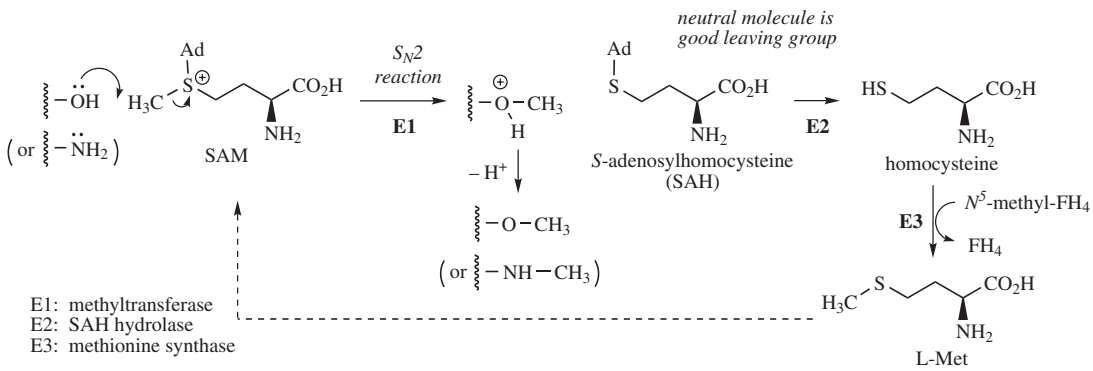
As indicated above, the C_5 isoprene unit in the form of **DMAPP** can be used to alkylate a nucleophile. In the elaboration of terpenoids and steroids, two or more C_5 units are joined together and the reactions are rationalized

Alkylation reactions: nucleophilic substitution

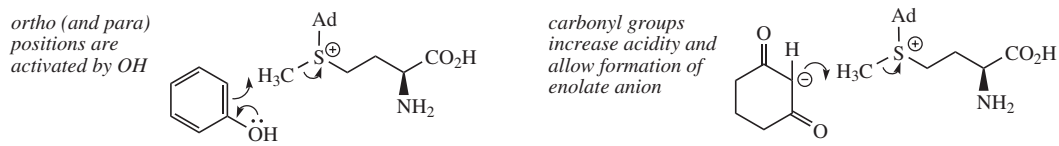
(a) formation of SAM



(b) *O*- and *N*-alkylation using SAM; regeneration of methionine



(c) *C*-alkylation using SAM



(d) *O*-alkylation using DMAPP

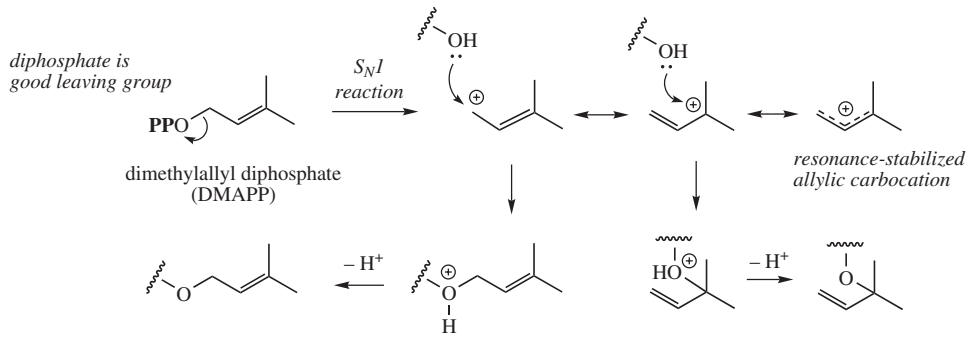


Figure 2.4

in terms of carbocation chemistry, including electrophilic addition of carbocations onto alkenes. DMAPP may ionize to generate a resonance-stabilized allylic carbocation as shown in Figure 2.4(d), and this can then react with an alkene, e.g. **isopentenyl diphosphate** (**IPP**), as depicted in Figure 2.5(a). The resultant carbocation may then lose a proton to give the uncharged product geranyl diphosphate (GPP). Where the alkene and carbocation functions reside in the same molecule, this type of mechanism can also be responsible for cyclization reactions [Figure 2.5(a)].

The initial carbocation may be generated by a number of mechanisms, important examples being loss of a leaving group, especially diphosphate (i.e. S_N1-type ionization), protonation of an alkene, and protonation/ring opening of epoxides [Figure 2.5(b)]. **SAM** may also alkylate alkenes by an electrophilic addition mechanism, adding a C_1 unit, and generating an intermediate carbocation; this is simply an extension of the protonation reaction.

The final carbocation may be discharged by loss of a proton (giving an alkene or sometimes a cyclopropane ring) or by quenching with a suitable nucleophile, especially water [Figure 2.5(c)].

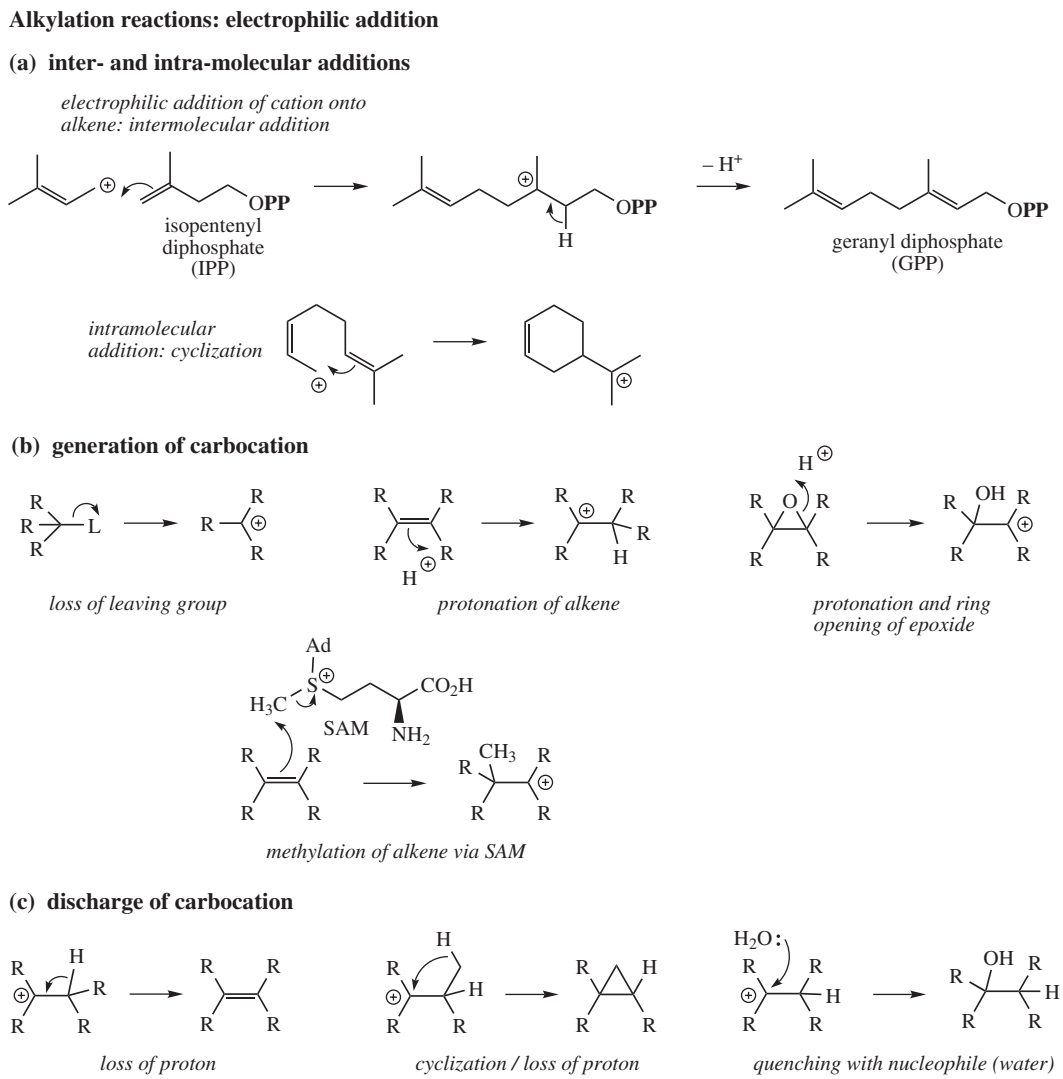


Figure 2.5

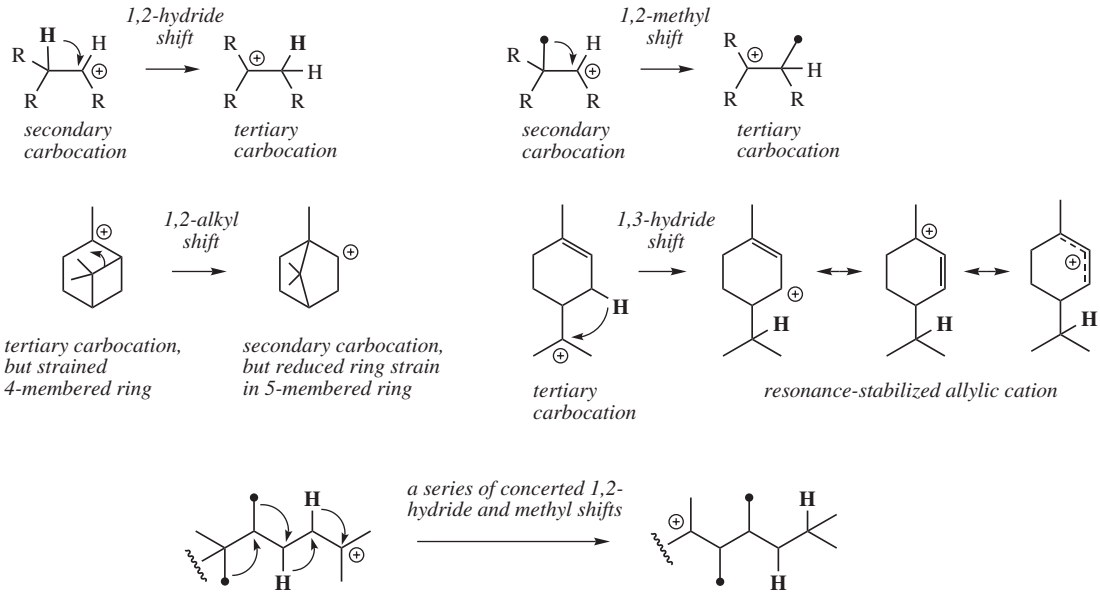


Figure 2.6

Wagner–Meerwein Rearrangements

A wide range of structures encountered in natural terpenoid and steroid derivatives can only be rationalized as originating from C_5 isoprene units if some fundamental rearrangement process has occurred during biosynthesis. These rearrangements have, in many cases, been confirmed experimentally, and are almost always consistent with the participation of carbocation intermediates. Rearrangements in chemical reactions involving carbocation intermediates, e.g. S_N1 and E1 reactions, are not uncommon and typically consist of 1,2-shifts of hydride, methyl, or alkyl groups. Occasionally, 1,3- or longer shifts are encountered. These shifts, termed **Wagner–Meerwein rearrangements**, are readily rationalized in terms of generating a more stable carbocation or relaxing ring strain (Figure 2.6). Thus, tertiary carbocations are favoured over secondary carbocations, and the usual objective in these rearrangements is to achieve tertiary status at the positive centre. However, a tertiary to secondary transition might be favoured if the rearrangement allows a significant release of ring strain. These general concepts are occasionally ignored by nature, but it must be remembered that the reactions are enzyme mediated and that carbocations may not exist as discrete species in the transformations. An interesting feature of some biosynthetic pathways, e.g. that leading to steroids, is a remarkable series of concerted 1,2-migrations rationalized via carbocation chemistry, but entirely a consequence of the enzyme's participation (Figure 2.6).

Aldol and Claisen Reactions

The **aldol** and **Claisen** reactions both achieve carbon–carbon bond formation; in typical base-catalysed chemical reactions, this depends upon the generation of a resonance-stabilized enolate anion from a suitable carbonyl system (Figure 2.7). Whether an aldol-type or Claisen-type product is formed depends on the nature of X and its potential as a leaving group in the alkoxide anion intermediate. Thus, chemically, two molecules of acetaldehyde yield aldol, whilst two molecules of ethyl acetate can give ethyl acetoacetate. These processes are vitally important in biochemistry for the elaboration of both secondary and primary metabolites, but the enzyme catalysis obviates the need for strong bases, and probably means the enolate anion has little more than transitory existence. Nevertheless, the reactions do appear to parallel enolate anion chemistry and are frequently responsible for joining together of C_2 acetate groups.

In most cases, the biological reactions involve coenzyme A esters, e.g. **acetyl-CoA** (Figure 2.8). This is a

Aldol and Claisen reactions

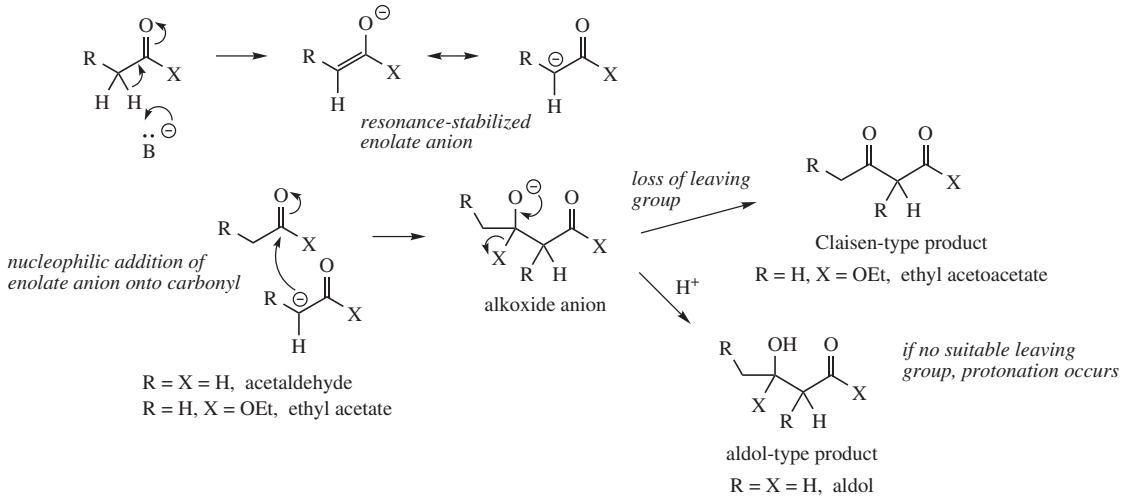


Figure 2.7

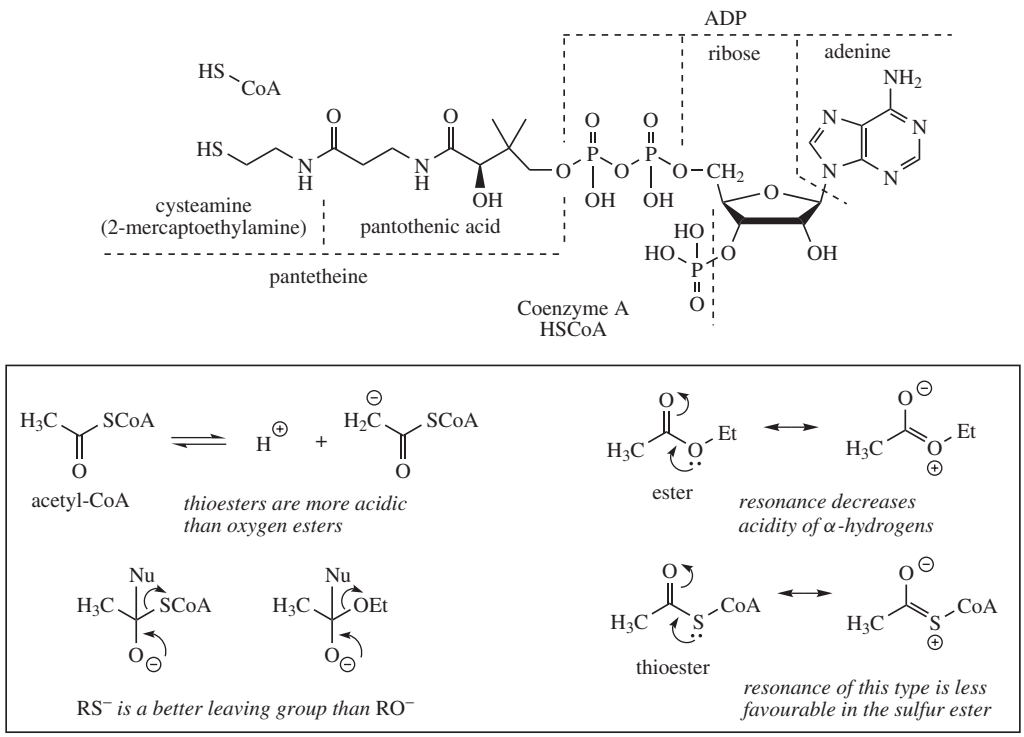


Figure 2.8

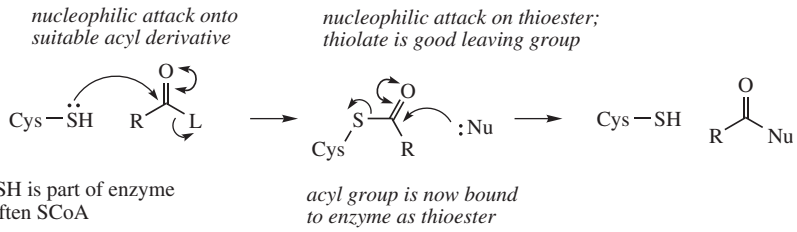


Figure 2.8 *(Continued)*

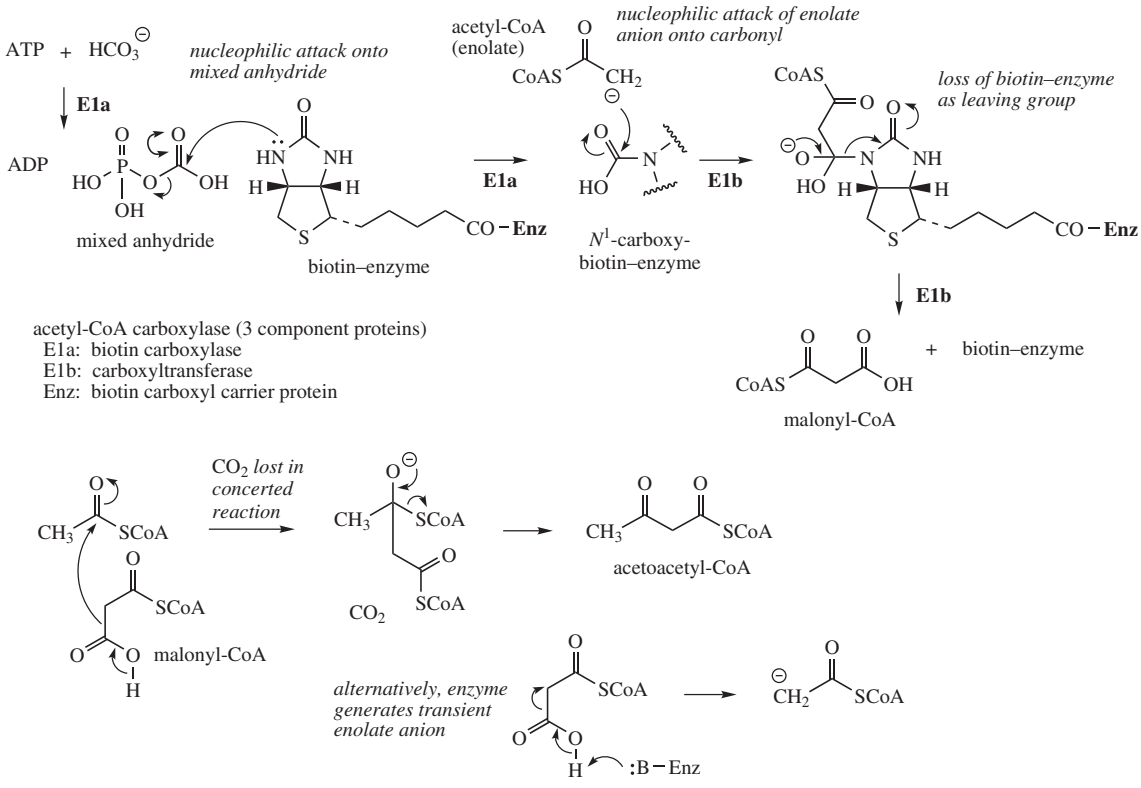


Figure 2.9

thioester of acetic acid, and it has significant advantages over oxygen esters, e.g. ethyl acetate, for two main reasons. First, the α-methylene hydrogen atoms are now more acidic, comparable in fact to those in the equivalent ketone, and this increases the likelihood of generating an enolate anion. This is explained in terms of electron delocalization in an ester function (Figure 2.8). This type of delocalization is more prominent in the oxygen ester than in the sulfur ester, due to oxygen's smaller size and closer proximity of the lone pair for overlap with carbon's orbitals. Second, the thioester has a much more favourable leaving group than the oxygen ester. The combined effect is to increase the reactivity for both the aldol- and Claisen-type reactions. Thioester linkages also provide a means of covalently bonding suitable substrates to enzymes, prior to enzymic modification, followed by subsequent release (Figure 2.8).

Claisen reactions involving acetyl-CoA are made even more favourable by first converting acetyl-CoA into **malonyl-CoA** by a carboxylation reaction catalysed by the three-component enzyme acetyl-CoA carboxylase. The reaction requires CO_2, ATP and the coenzyme biotin (Figure 2.9). ATP and CO_2 (solubilized as bicarbonate, HCO_3^-) form the mixed anhydride, a reaction also

catalysed by the enzyme, and this is used to carboxylate the coenzyme which is bound in a biotin–enzyme complex. The carboxylation reaction is effectively a nucleophilic attack of the cyclic urea onto the mixed anhydride. Fixation of CO_2 by biotin–enzyme complexes is not unique to acetyl-CoA; another important example occurs in the generation of oxaloacetate from pyruvate in the synthesis of glucose from non-carbohydrate sources (gluconeogenesis). The conversion of acetyl-CoA into malonyl-CoA means the α-hydrogen atoms are now flanked by two carbonyl groups, and have increased acidity. Thus, a more favourable nucleophile is provided for the Claisen reaction. No acylated malonic acid derivatives are produced, and the carboxyl group introduced into malonyl-CoA is simultaneously lost by a decarboxylation reaction during the Claisen condensation (Figure 2.9). An alternative rationalization is that decarboxylation of the malonyl ester is used to generate the transient acetyl enolate anion without any requirement for a strong base. Thus, the product formed from acetyl-CoA as electrophile and malonyl-CoA as nucleophile is acetoacetyl-CoA, exactly the same as in the condensation of two molecules of acetyl-CoA. Accordingly, the role of the carboxylation step is clear-cut: the carboxyl activates the α-carbon to facilitate the Claisen condensation and it is immediately removed on completion of this task. By analogy, the chemical Claisen condensation using the enolate anion from diethyl malonate in Figure 2.10 proceeds much more favourably than that using the enolate anion from ethyl acetate. The same acetoacetic acid product can be formed in the malonate condensation by hydrolysis of the acylated malonate intermediate and decarboxylation of the *gem*-diacid.

Analogous carboxylations of some other thioesters may occur; for example, propionyl-CoA may be converted into methylmalonyl-CoA.

Both the **reverse aldol** and **reverse Claisen** reactions may be encountered in the modification of natural product molecules. Such reactions remove fragments from the basic skeleton already generated, but may extend the diversity of structures. The reverse Claisen reaction is a prominent feature of the **β-oxidation** sequence for the catabolic degradation of fatty acids (Figure 2.11): a C_2 unit as acetyl-CoA is cleaved off from a fatty acid chain, leaving it two carbon atoms shorter in length. Though the terminology β-oxidation strictly refers to the introduction of the new carbonyl group, it is usually understood to include the chain shortening.

Imine Formation and the Mannich Reaction

Formation of C–N bonds is frequently achieved by condensation reactions between amines and aldehydes or ketones. A typical nucleophilic addition is followed by elimination of water to give an **imine** or **Schiff base** [Figure 2.12(a)]. Of almost equal importance is the reversal of this process, i.e. the hydrolysis of imines to amines and aldehydes/ketones [Figure 2.12(b)]. The imine so produced, or more likely its protonated form the iminium cation, can then act as an electrophile in a **Mannich reaction** [Figure 2.12(c)]. The nucleophile might be provided by an enolate anion, or in many examples by a suitably activated centre in an aromatic ring system. The Mannich reaction is encountered throughout alkaloid biosynthesis, and in its most general form involves combination of an amine (primary or secondary), an aldehyde or ketone, and a nucleophilic carbon. Secondary amines will react with the carbonyl compound to give an iminium cation (quaternary Schiff base) directly; thus, the additional protonation step is not necessary.

It should be appreciated that the Mannich-like addition reaction in Figure 2.12(c) is little different from

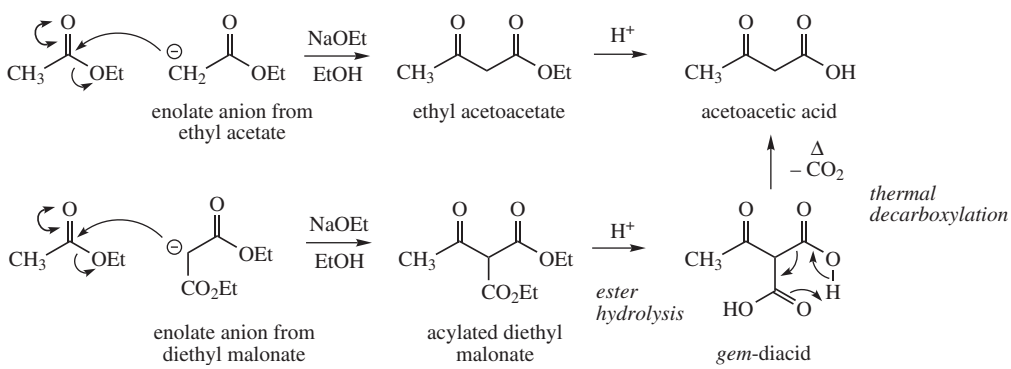


Figure 2.10