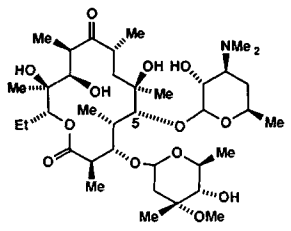
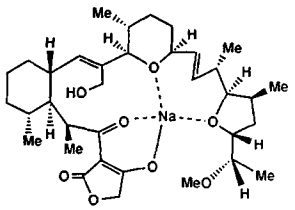
The Extraordinary Enzymes Involved in Erythromycin Biosynthesis

By *James Staunton**

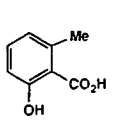
The polyketides form a family of compounds which have very diverse structures and no obvious structural relationship, at least on superficial examination. Representatives of some of the main classes are presented in Scheme 1. Many of these compounds are of great commercial interest. For example, erythromycin, tetronasin, and oxytetracycline are all antibiotics which are used in medical practice and in animal husbandry. Another polyketide metabolite, FK 506 (not shown), is currently causing great excitement for its ability to act as an immunosuppressant following organ transplant surgery.[**]



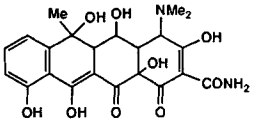
Erythromycin A



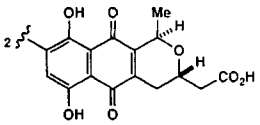
Tetronasin



6-MSA



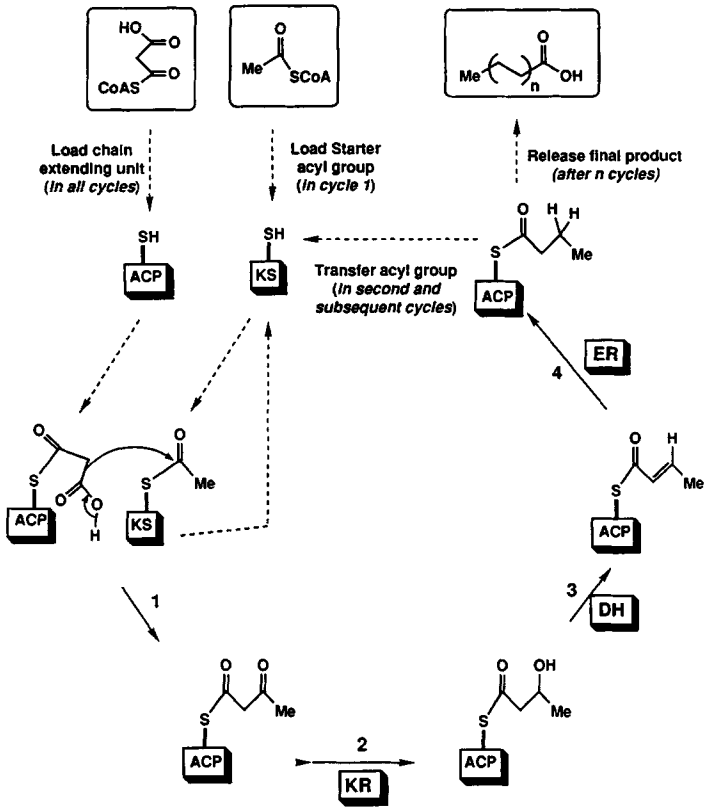
Oxytetracycline



Actinorhodin

Scheme 1. Types of polyketide metabolites. 6-MSA = 6-methylsalicylic acid.

Despite their structural diversity the polyketides are thought to be related in biosynthesis, at least in the early steps of the pathways in which the essential core of the structure is assembled. This assembly process is thought to involve a variation of the standard biosynthetic cycle by which fatty acids are produced. In this cycle the growing fatty acyl chain is extended by the repeated addition of two methylene groups.[1] A representative reaction sequence is shown in Scheme 2. Initially, acetate is condensed with malonate to give acetoacetate with simultaneous loss of CO_2. Steps 2, 3, and 4 follow an unexceptional strategy to convert the ketone carbonyl to a methylene. The resulting extended acyl group then reenters at the starting point, and the cycle is repeated until the appropriate length of chain is reached. A special characteristic of these systems is that the growing chain is bonded via a thioester link to a protein thiol group at all stages, until the final product is released. The system of enzymes which carries out these steps is known as the fatty acid synthase (FAS) system.

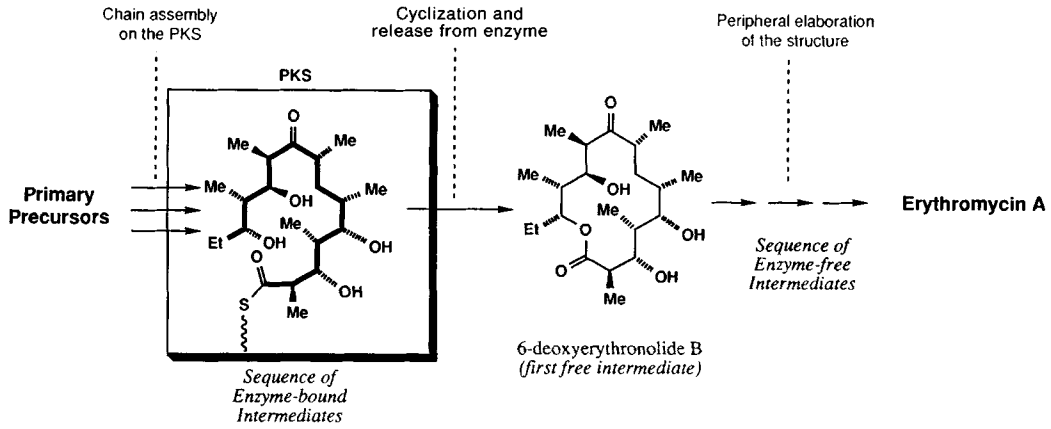


Scheme 2. Reactions and enzymes of fatty acid biosynthesis. KS = ketoacyl-synthetase; KR = ketoreductase; DH = dehydratase; ER = enoylreductase; ACP = acyl-carrier protein; CoA = Coenzyme A.

On casual inspection it may seem improbable that any of the polyketides of Scheme 1 is related to the simple fatty acids, but ample evidence has been accumulated to support this proposition. Studies with mutants have provided important though indirect evidence. When a mutant of a polyketide-producing organism lacks the ability to produce one of the biosynthetic enzymes, the preceding intermediate may accumulate in sufficient quantities for it to be isolated and identified. Such experiments with the erythromycin producer, *Saccharopolyspora erythraea* (formerly *Streptomyces erythraea*), led to the identification of several late intermediates of the biosynthetic pathway.[2] These compounds lacked some of the features of the erythromycin A structure, but nothing simpler than the macrocyclic lactone 6-deoxyerythronolide B was discovered. These experiments therefore indicate that the biosynthesis of erythromycin takes place in two phases, as is shown in Scheme 3, linked by the pivotal first enzyme-free intermediate, 6-deoxyerythronolide B. Al-

[*] Dr. J. Staunton
University Chemical Laboratory
Lensfield Road, GB-Cambridge CBZ 1EW (UK)

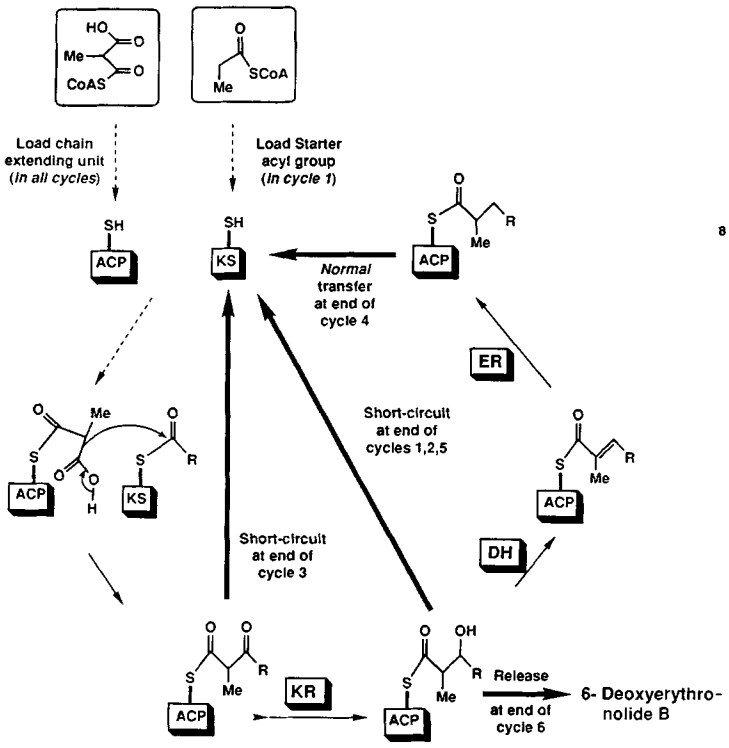
[**] See the Highlight in Number 8/1991: H. Kessler, D. F. Mierke, D. Donald, M. Furber, *Angew. Chem. 103* (1991) 968; *Angew. Chem. Int. Ed. Engl. 30* (1991) 954.



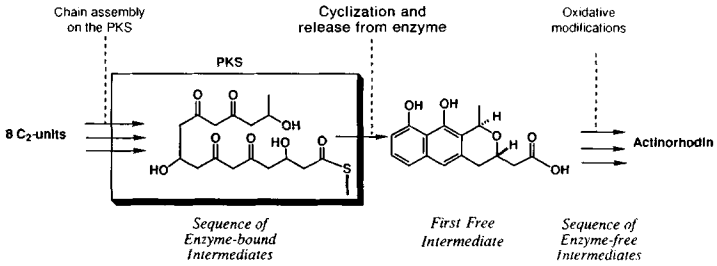
Scheme 3. Overview of the erythromycin A biosynthetic pathway. PKS = polyketide synthetase system.

though they throw no light on the mode of assembly of the key macrolide core of the molecule, these results support the idea that the early intermediates remain enzyme-bond, and so cannot accumulate in sufficient quantities for detection. By analogy with FAS systems, the enzymes involved in the chain assembly stage of a polyketide biosynthetic pathway are collectively known as the polyketide synthase (PKS) system.

The structure of the proposed final intermediate bound to the PKS in Scheme 3 is highlighted with heavy lines to emphasize that it contains a chain of carbons terminating in a carboxyl residue, and so has some structural resemblance to a simple fatty acid. Greater structural complexity is conferred on the antibiotic precursor by a series of branching methyl groups and by oxygen substituents. Both of these structural features can easily be introduced by a variation of the fatty acid biosynthetic cycle. The use of propionate rather than acetate as the chain starter acyl, and methyl malonate instead of malonate for the chain extension steps in Scheme 2 results in the formation of a carbon skeleton with methyl branching. Hydroxyl or ketone substituents could be left at the appropriate sites by omitting some or all of the steps normally used to convert each new ketone group to a methylene. In the case of 6-deoxyerythronolide B, the sequence of reactions catalyzed by the PKS would then take the form outlined in Scheme 4. The heavy arrows indicate where acyl groups need to be transferred in the various cycles, in order to leave the appropriate array of oxygen functions in the final enzyme-bound intermediate. Subsequent lactonization, perhaps catalyzed by a component of PKS, would then lead to the release of 6-deoxyerythronolide B. There are many classical biosynthetic incorporation studies using simple primary precursors variously labeled with isotopes of carbon, hydrogen, and oxygen in support of this scheme, and more recent incorporation studies with 'partly formed' chains, corresponding to enzyme-bound intermediates, have provided added insights.[3]



Scheme 4. Proposed mode of operation of the erythromycin PKS.



Scheme 5. Proposed biosynthesis pathway for actinorhodin.

Encouraged by progress made in studies of the genetics of the actinorhodin pathway by the Hopwood group in Norwich, many groups engaged in studies of polyketide biosynthesis have recently turned to a genetic approach in an effort to gain a clearer understanding of the molecular basis of chain assembly.[4] Actinorhodin has a structure consistent with a derivation from C_2-units via the enzyme-bound polyketide intermediate shown in Scheme 5. Since most of the ketone residues of this chain have been left unmodified by PKS, subsequent cyclization and dehydration reactions could take place to form the second intermediate shown. The

proposed pathway leads to actinorhodin by further oxidative steps involving other enzymes.

Many genes coding for proteins involved in this biosynthesis pathway have now been identified. Of particular interest here are a set of four which code for proteins having a strong sequence homology with protein components of various fatty acid synthases, and a fifth which is thought to be the cyclase responsible for the process of aromatization. These proteins are very probably the components of PKS. Interestingly, the individual genes are discrete and therefore the corresponding enzyme activities are not covalently linked to each other as they are formed. The actinorhodin PKS therefore probably resembles the FAS systems of most bacteria in consisting of a dissociable set of cooperating enzymes. Several other PKS gene clusters associated with the generation of polyacetate aromatic compounds in various *Streptomyces* species show a similar organization.[4] In contrast, the PKS for biosynthesis of 6-methylsalicylic acid (6-MSA), which has been isolated, is composed of a similar set of enzymes activities, linked covalently to form a large multifunctional protein.[5] The active form of the PKS is formed by the aggregation of four such multifunctional proteins. This more elaborate molecular architecture, which has precedents in the FAS systems of yeasts, mammals, and certain bacteria, probably leads to a more efficient and possibly more robust catalytic apparatus. In all of these polyacetate PKS systems, only one gene codes for each catalytic activity in the FAS cycle, which means that individual active sites in the PKS complex participate in successive cycles of chain extension.

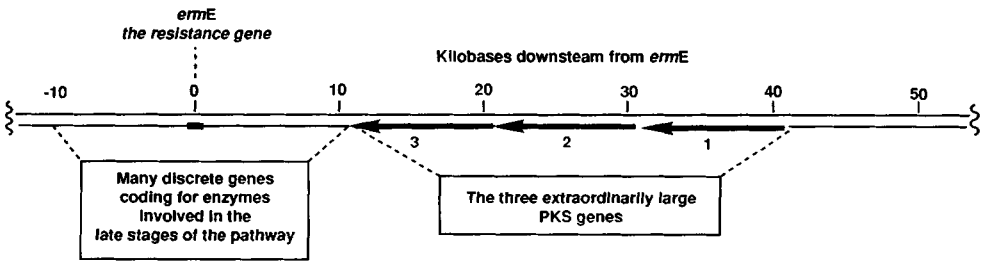
The breakthrough in locating the actinorhodin PKS genes has stimulated a search for PKS genes in other organisms. In one approach genetic probes based on the actinorhodin genes were used with considerable success to locate the PKS genes for various aromatic polyacetate metabolites.[4] A second approach, also based on the pioneering work of the Norwich group, proved successful with the erythromycin genes, although the size of the task proved much greater than anyone imagined at the outset.[6, 7] Firstly, the gene coding for erythromycin resistance in the producing organism was located. By working initially on the assumption that the biosynthetic genes would be clustered in the vicinity of the resistance gene, as had been observed for the actinorhodin genes, the neighboring DNA was sequenced in a search for genes coding for FAS type proteins. This search was encouraged by the finding stemming from the Abbott laboratories in Chicago that mutations that prevented synthesis of 6-deoxyerythronolide B could all be localized to near the resistance gene.

The resulting picture which has emerged for the gene structure of the enzyme system for erythromycin biosynthesis is outlined in Scheme 6. On both sides of the resistance gene there are regions containing about 10000 pairs (10 kbp) of DNA which contain genes thought to code for proteins involved in the late stages of the pathway, in which 6-deoxyerythronolide B is converted into the erythromycin A. In confirmation, deletion of individual genes in these regions produced mutant organisms from which late intermediates have been isolated.

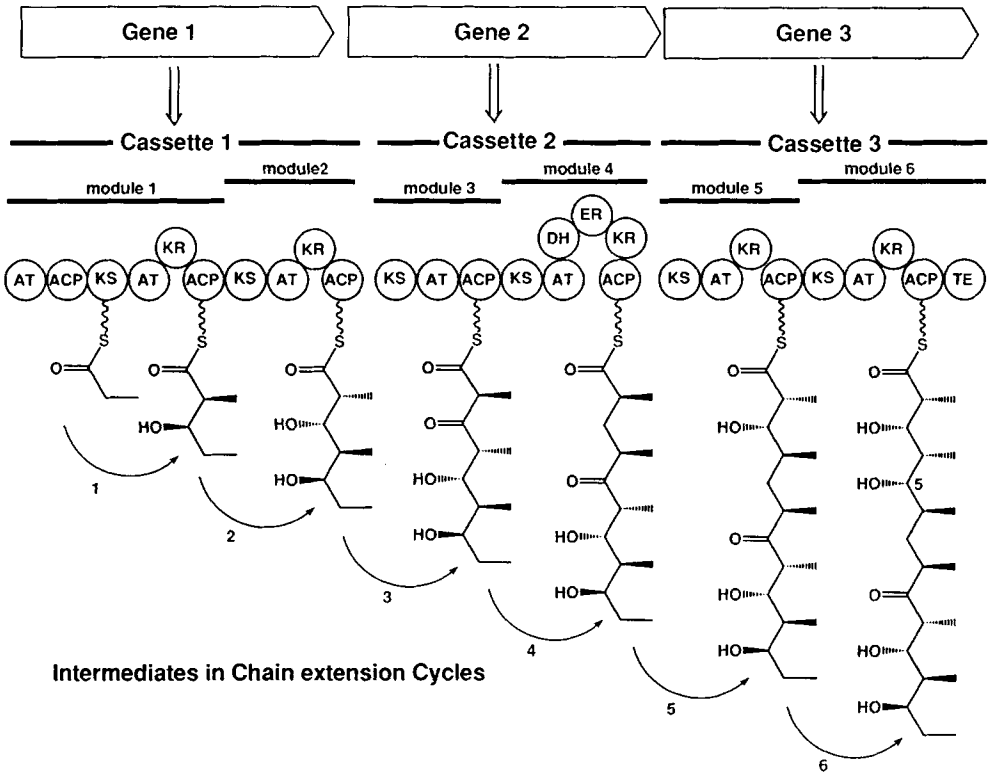
Sequencing further away from the resistance gene has subsequently revealed a region of DNA which contains sections corresponding to the elusive PKS gene cluster (Scheme 6). It became clear that the erythromycin PKS is much more elaborate than that associated with the biosynthesis of actinorhodin and other polyacetate aromatic compounds. The first indication of this extra complexity came with the discovery that there are several different genes coding for specific components of the PKS (e.g., there is more than one acyl-carrier protein (ACP) gene). This suggested that different cycles of chain extension might employ different versions of specific component enzymes.

What is perhaps even more surprising is the complexity of the organization of the enzyme activities. This was first revealed when a group from Cambridge published the complete sequence of one of the open reading frames (i.e., a potential gene) coding for part of the erythromycin PKS.[6] This exceptionally large open reading frame, containing more than 10 kbp of DNA, codes for a single giant protein of molecular weight 332472 Da, particular segments of which show sequence homology with individual FAS proteins. In all there are eight catalytic activities, corresponding to two different sets of the enzymes required for chain extension, together with a thioesterase.

More recently the complete sequence of the erythromycin PKS genes has been published by *Katz* and co-workers from the Abbott Laboratories.[7] It consists of three large open reading frames each coding for a large multifunctional protein having sufficient catalytic sites to carry out two chain extension cycles. It would appear, therefore, that there is an individual enzyme for each and every step carried out by the erythromycin PKS. The activities seem to be organized in six "modules", all of which contain the active sites for a condensation step, and all but one contain some or all of the active sites associated with conversion of the ketone group into a methylene in the fatty acid biosynthetic cycle. These modules are further linked in pairs, for which I suggest the term cassette. Three such cassettes collaborate in some undefined way to produce the first enzyme-free-intermediate, 6-de-



Scheme 6. Outline map of the cluster of genes coding for enzymes involved in erythromycin biosynthesis.



Scheme 7. The erythromycin polyketide synthetase system (PKS): primary organization of the genes and their corresponding protein cassettes. Examples of intermediates in the chain extension cycles are given. However, see caveat in text.

oxyerythronolide B, from one unit of propionate and six of methylmalonate.

Scheme 7 shows in more detail how the PKS genes are arranged on the genome, and how the catalytic sites are ordered in the primary structure of the protein chain in each corresponding cassette. A ligand is shown attached to each module to indicate how the modules may operate in assembling the polyketide chain. It should be stressed, however, that there are many possible ways for the cassettes to fold, possibly bringing complementary activities sited in different modules close together. In the functioning of the cassettes, therefore, the various activities which cooperate in a particular chain extension cycle may be drawn from different modules. Consequently, individual active sites with a particular function (e.g., the ACP's) need not necessarily come into play in the order in which they appear in the primary sequence of the protein, as is assumed in Scheme 7. At this time only one catalytic site, the ketone reductase (KR) site in module 5, has been firmly linked to a specific role in the elaboration of the erythromycin chain.[7] This was achieved by selective disruption of the DNA coding for this activity. The resulting mutant organism generated, instead of erythromycin, an analogue carrying a keto group at C5 of the macrolide ring. It is safe to conclude, therefore, that this reductase operates in the fifth cycle of chain extension, but not that its nearest neighbors in the primary sequence of the protein are its functioning partners.

In its size and sheer complexity this extraordinary PKS system breaks new ground, although in passing it should be mentioned that giant multifunctional proteins have also been implicated in the nonribosomal biosynthesis of some peptide secondary metabolites, and further examples may emerge as efforts to identify enzymes involved in other pathways of secondary metabolism gather pace.[8] The information obtained from sequencing the genes gives a fascinating insight into the primary organization of the PKS proteins, and some indication of how they function. An enormous multidisciplinary effort is required, however, to uncover fully the molecular basis of their action, and this will have to include direct studies of the proteins themselves. Encouraging progress on the isolation of PKS proteins from the erythromycin PKS has been reported from Cambridge.[9] A section of the open reading frame for cassette 3, corresponding to the thioesterase (TE) and ACP activities at the C-terminus, has been expressed at high levels in *Escherichia coli*. The resulting protein had the expected molecular mass (38019 Da), according to the electrospray mass spectrum, and it showed an ability to bind a standard inhibitor (phenylmethylsulphonyl fluoride) in a manner consistent with its expected thioesterase activity. More recently, the complete open reading frame coding for cassette 3 has been expressed in *E. coli* as a single protein with a molecular weight in the required range (ca 300 kDa by electrophoresis), confirming that these large genes are capable of expression as large single proteins.[10] Now that PKS proteins are becoming available, there is the exciting prospect that the activity of the constituent active sites may be demonstrated in vitro. The Cambridge group has already succeeded with a FAS ACP derived from the erythromycin producer, and in doing so has developed methodology appropriate to studying the PKS system.[11]

What other developments should we look for in the near future? It will be interesting to see if the organization of the erythromycin PKS as a series of modules arranged in cas-

settes holds good for other macrolide PKS's, and also possibly for ionophore-producing PKS's. This ordering of proteins in cassettes, which need not be exclusively bimodular, may hold the key to the precise control exercised by the synthase as a whole in carrying out its synthetic operations. It would, for example, help rationalize both the early observations of *Celmer* et al. on the common patterns of structure and stereochemistry which run through families of macrolides,[12, 13] and more recent related analyses of ionophore structures.[14] Perhaps regions with common structure in diverse metabolites are produced in different organisms by closely related cassettes derived from a common ancestral protein. Looking ahead, it may be possible to engineer organisms to produce novel chimeric metabolites by switching genes coding for specific cassettes from one producer to another.

If substantial amounts of active PKS proteins can be isolated by overexpression, it will be possible to investigate the molecular mechanisms by which the cassettes carry out their specific tasks. The protein–protein interactions which must play a major part in the collaboration between cassettes will make a fascinating study. Equally interesting are the molecular interactions which occur between protein components and substrates, which presumably help define the substrate specificity of the enzymic reactions. Progress in such studies could be of particular significance to synthetic chemists who have made great efforts in recent years to develop new methodology for macrolide synthesis.[15] Isolated PKS cassettes used in vitro may well prove to be convenient and efficient alternative synthetic tools for carrying out limited transformations on natural substrates and substrate analogues. In the longer term, it may be feasible to employ different combinations of cassettes in vitro, to generate complex chimeric products with greater control than would be possible in vivo. Chemists as well as biologists have reason to watch developments in this field.

German version: *Angew. Chem. 103* (1991) 1331

[1] S. Numa (Ed.): *Fatty Acid Metabolism and its Regulation* (in *Comprehensive Biochemistry, Vol. 7*) Elsevier, Amsterdam 1984.
[2] E. T. Seno, C. R. Hutchinson in S. W. Queener, I. E. Day (Eds.): *The Bacteria, Vol. 9*, Academic Press, New York p. 231.
[3] D. E. Cane, C. Yang, *J. Am. Chem. Soc. 109* (1987) 1255.
[4] D. A. Hopwood, D. H. Sherman, *Annu. Rev. Genet. 24* (1990) 37.
[5] J. Beck, S. Ripka, A. Siegner, E. Schiltz, E. Schweizer, *Eur. J. Biochem. 192* (1990) 487.
[6] J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature 348* (1990) 176.
[7] S. Donadio, M. J. Staver, J. B. McAlpine, J. B. Swanson, L. Katz, *Science 252* (1991) 675.
[8] H. Kleinkauf, H. von Doehren, *FEBS Lett. 268* (1990) 405.
[9] P. Caffrey, B. Green, L. C. Packman, B. J. Rawlings, J. Staunton, P. F. Leadlay, *Eur. J. Biochem. 195* (1991) 823.
[10] G. A. Roberts, personal communication.
[11] A. M. Bridges, P. F. Leadlay, W. P. Revill, J. Staunton, *J. Chem. Soc. Chem. Commun. 1991*, 778.
[12] W. D. Celmer, *J. Am. Chem. Soc. 87* (1965) 1801.
[13] D. E. Cane, W. D. Celmer, J. W. Westley, *J. Am. Chem. Soc. 105* (1983) 3594.
[14] D. O'Hagan, *Nat. Prod. Rep. 1989*, 205.
[15] Editorial comment: see also the Highlight by *J. Mulzer* in the November issue.

$LiClO_4$ in Ether—an Unusual Solvent

By *Herbert Waldmann**

Many reactions can be influenced in a variety of ways by the solvent employed. This is especially the case when polarized transition states or ionic intermediates are involved in the reaction and when the solvent is nucleophilic or electrophilic. A case to the contrary is the Diels-Alder reaction, which remains largely unaffected by the surrounding organic medium. In the mid-eighties, however, *Breslow* et al.[1] and *Grieco* et al.[2] demonstrated that Diels-Alder reactions proceed with increased reaction rate and with improved *endo-exo* selectivity when they are carried out, not in organic solvents but in aqueous solutions. The effect is further enhanced by salts such as LiCl (salting-in effect), whereas the addition of guanidinium chloride has the opposite effect (salting-out effect). The use of water as solvent for such cycloadditions had already been described earlier by *Alder* et al.[3a] and later by *Koch* et al.[3b] The accelerating effect of this reaction medium is also manifested in many other reactions,[4] e.g. asymmetric hetero-Diels-Alder reactions[4b] and asymmetric nonhetero-Diels-Alder reactions,[4c, d] nucleophilic additions to iminium ions[4e] and carbonyl compounds,[4f] Claisen rearrangements,[4g] the benzoin condensation,[1b] and aldol reactions.[4h] It is attributed to the fact that a suitable aggregation is generated by hydrophobic interactions between the reaction partners (hydrophobic effect), and so exercises an "internal pressure" on the reactants encapsulated in "solvent cavities" whose effects are, in turn, comparable with a high external pressure, at least in the case of the Diels-Alder reaction.

Last year, *Grieco* et al.[5] described a solvent system, namely a 5 M solution of $LiClO_4$ in diethyl ether, which has a comparable, if not greater accelerating effect on Diels-Alder reactions. Already in 1986 *Sauer* et al.[6] investigated the steric course of Diels-Alder reactions in such solvent systems in order to determine the polarity of ethereal $LiClO_4$ solutions, and had already recommended them as polar solvents for organic chemical reactions. *Grieco* et al. then demonstrated the astounding properties of these solutions (see Scheme 1), which according to *Pocker* et al.[7] must be regarded as mixtures of the mono- and diether adduct of $LiClO_4$. The reaction of cyclopentadiene **1** with ethyl acrylate **2**

[*] Prof. Dr. H. Waldmann
Institut für Organische Chemie und Biochemie der Universität
Gerhard-Domagk-Strasse 1
W-5300 Bonn (FRG)