Cationic Polyene Cyclizations. A Computer Assisted Synthesis Approach

Jan-Willem Boiten and Jan H. Noordik*

CAOS/CAMM Center, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Marinus B. Groen

Organon Scientific Development Group, P.O. Box 20, 5340 BH Oss, The Netherlands

Received April 21, 1993®

The implementation of cationic polyene cyclizations in a synthesis planning program requires a special approach which focuses primarily on the synthetically useful cyclizations. This paper describes such an approach which is based on the recognition of strategic disconnections in the skeleton of a target molecule. This approach has been implemented in the LHASA system. The output of this implementation is illustrated by a LHASA analysis of two synthetic target structures.

INTRODUCTION

The enzymatic conversion of Squalene to Lanosterol[1] (Figure 1) is a key step in the biosynthesis of steroids. In a single step, a number of rings are stereoselectively formed from an acyclic chain of olefinic bonds. This beautiful enzymatic reaction has inspired chemists to their biomimetic approach to the total synthesis of polycyclic natural products. Figure 2 shows an example of such a polyene cyclization.[2] These cyclizations have been extensively studied over the past 25 years, resulting in a number of useful synthetic methods.[3-7] A major topic of discussion concerns the mechanism of the reaction;[6-8] opinions vary from a more stepwise mechanism[9,10] to a purely concerted reaction.[11] Although the concertedness of the reaction is still open to debate, it is useful as a thought process to treat the reaction as a stepwise one. Such a stepwise mechanism may be divided into three stages: the initiation, one or more ring closure steps, and the termination. The initiation involves the formation of a cation from a suitable functionality, which starts the chain reaction. This chain reaction can propagate over several ring closure steps, but finally a cation will remain which must be trapped to terminate the polyene cyclization properly. This thought process can be mapped very efficiently to a computer implementation, given a suitable environment to perceive structures, to handle reaction sequences, and to generate products. Such an environment is presented by the LHASA system,[12] and our approach of cationic polyene cyclizations has been developed within the framework of this system.

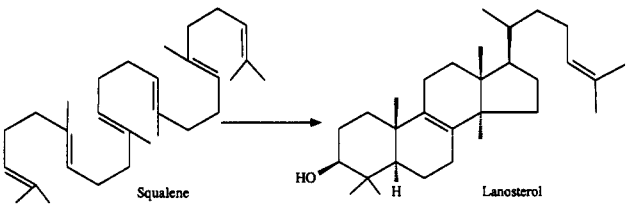

Figure 1. Biosynthesis of Lanosterol from Squalene.

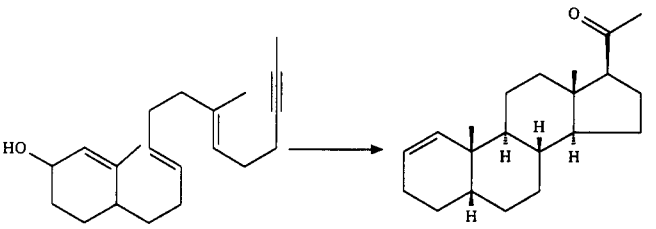

Figure 2. Example of a polyene cyclization.

The LHASA System. Over the past 20 years, several computer programs have been developed to assist the chemist with the synthesis of complex organic molecules.[13,14] One of the major systems in this field is the LHASA program,[12,15,16] which has been under continuous development at Harvard University since the early 1970s and more recently also at some other groups in several countries. The program evolved from E. J. Corey's ideas about retrosynthetic synthesis planning,[17-19] and its main feature is the analysis of a target molecule in a rigorously retrosynthetic fashion, thus generating a number of retroreactions. The starting materials of these retroreactions ("precursors") are usually not directly available from the chemical suppliers but require further analysis instead. The user selects the precursors which look promising and submits them for further analysis. In this way, user and program interactively prepare a synthesis design. The complete analysis is summarized in a tree-shaped diagram (the retrosynthetic tree) to keep track of the synthesis scheme.

LHASA is an "expert system", consisting of two parts: a controlling program, mainly written in FORTRAN, and a chemical knowledge base written in a "chemical English" language (CHMTRN). The knowledge base consists of a large number of retrosynthetic reaction definitions called transforms. A transform is not a compilation of literature precedents of a reaction, but a program-type reaction description in which the scope and limitations of the reaction have been translated into empirical rules written in CHMTRN. The special-purpose language CHMTRN includes many chemical terms, as well as the usual programming constructions like block-ifs and subroutines. These special features of CHMTRN make it an excellent tool for the input of chemical knowledge.

It is impossible to apply each theoretically possible transform for a given target molecule, as the tree would explode to numerous precursors within a few cycles of the analysis (combinatorial explosion). Apparently, some kind of a strategic screen has to be applied to cut down the number of branches in the retrosynthetic tree. The concept of simplifying transforms is used for this purpose, which implies that LHASA will only use transforms which simplify the target molecule. Transforms which result in a precursor that is more complicated to synthesize than the target molecule need not to be

* Abstract published in *Advance ACS Abstracts*, September 1, 1993.

0095-2338/93/1633-0727$04.00/0

considered by the program. The main shortcoming of this approach is the lack of a clear-cut definition of the concept of simplification. LHASA defines it as a decrease in cyclic connectivity, removal of appendages, generation of easily-synthesized substructures and/or functional groups, disconnection of long chains, removal of stereocenters, etc. Within this concept, still many questionable and nonstrategic transforms will be applied, exposing one of the main objectives of the program; that is, it should be a generator of synthetic ideas. This aim implies that questionable reactions get the benefit of the doubt. Within the interactive setup of LHASA, this approach does not have to endanger the synthetic design process, because the user can discontinue most branches in the retrosynthetic tree, focusing on the most promising (retrosynthetic) reaction sequences only. In addition to the transforms which meet the goal of simplification, LHASA also contains a large number of nonsimplifying transforms. These so-called subgoal transforms[20] perform mainly functional group interchanges and additions to pave the road for a goal transform. Using subgoal transforms, LHASA is able to produce short reaction sequences without any user interference. However, an unlimited usage of subgoal transforms might again lead to an unmanageably large number of precursors. Therefore, multistep subgoal usage is restricted during normal use of the program, leading to the need for a special device to create far-reaching subgoal sequences for some powerful reactions which would be suppressed otherwise. This special feature, called long-range transform,[21-23] directs the subgoal sequences which eventually lead to the display of a powerful reaction. Typical examples of long-range transforms are the Diels–Alder reaction[23] and the Robinson annulation.[21] This paper describes the development and operation of a new long-range transform dedicated to cationic polyene cyclizations.

DEVELOPMENT OF A NEW LONG-RANGE TRANSFORM: OVERVIEW

Polyene Cyclizations: The Retron Problem. The efficiency of an interactive synthesis planning system like LHASA is fully lost, if there is no way to focus quickly on the few useful transforms and to reject the majority of transforms which are way out of any use in the synthesis of the target. For this purpose, a special substructure, the retron, has been defined for each transform in the knowledge base. LHASA will only contemplate a transform if the retron of that transform can be identified as a substructure of the target; i.e., the retron is the key to enter the heuristics of the transform. However, the wide range of polycyclic structures which can be synthesized using cationic polyene cyclizations seriously obstructs the definition of the retron for this type of reaction. The retron approach necessitates the definition of a different retron for each ring system accessible by a polyene cyclization. This problem, inherent to the traditional LHASA approach to polyene cyclizations, has already been recognized in the early days of LHASA development and has been "solved" by a workaround. Instead of treating a polyene cyclization as a single step reaction, the reaction was treated as consisting of initiation, propagation, and termination steps and each step was described in a separate transform. Consequently, instead of having to develop a transform for each thinkable ring system, each with its own retron, only a limited number of transforms were coded in the LHASA knowledge base.

On first sight, this stepwise approach offers a perfectly reasonable solution to the problem, but a closer look will change this view, as several serious shortcomings will become apparent. We performed a detailed LHASA analysis of the synthesis of several steroid skeletons which indisputably showed that it was impossible to create a polyene cyclization sequence in LHASA, without a more than average knowledge of this reaction type. Transforms showing termination steps came up each time a suitable termination group was found in the target resulting in a cationic precursor (the stepwise analysis of a polyene cyclization will necessarily begin from the termination site due to the retrosynthetic character of the program). The next step in the cyclization is displayed only if the user selects the precursor resulting from a termination step for further analysis, and the desired propagation steps will be found among many other retroreactions. Subsequently, all resulting precursors must be reexamined, and this process must be repeated until the final initiation step has been reached. This method has the serious drawback that many user choices are involved. After displaying one step in the cyclization, the user must select a precursor for further analysis, but usually LHASA will have produced a whole series of precursors. We found in our analyses of steroid molecules typically 10–15 precursors resulting from termination steps. The selection of the precursor which eventually will lead to a complete polyene cyclization, requires a thorough knowledge of this type of reaction. Consequently, the user will only find a useful polyene cyclization if he is familiar with this method; certainly not a situation in which one might choose to use a synthesis planning system which should be a generator of synthetic ideas. A second and related objection against the approach to polyene cyclizations consisting of three isolated steps is the incompatibility with subgoal transforms. The use of subgoals in a stepwise approach significantly increases the selection problem described above, as it increases the number of termination steps shown to the user.

This discussion of the main flaws of the stepwise approach to polyene cyclization leads to a picture of the requirements for a new approach to these cyclizations. Apparently, it is insufficient to treat the cyclization as a chain of single reaction steps, so it will be necessary to consider the cyclization as an entity. Without any preselection, this would lead to a huge amount of (retrosynthetic) cyclizations, which are theoretically possible but practically useless, as they lead to inaccessible precursors. As a consequence, a strict selection must be made in the early stages of the analysis, to avoid this explosion of the number of cyclizations. Furthermore, the new LHASA approach must meet the requirements of a high-quality transform: Firstly, it must contain a thorough description of the scope and limitations of the reaction involved. This requirement demands an accurate description of allowed and disallowed features, particularly at the initiation and termination sites. Secondly, it must be able to efficiently use the subgoal power of LHASA to remove unfavorable features and introduce desired substructures. To meet these requirements, a completely new module of the LHASA program has been developed, in which the obvious construction of the long-range transform, offering a tremendous subgoal power, was used. This subgoal power must, however, be directed to the useful cyclizations and avoid the large number of theoretically possible, but practically useless, ones. Our approach for meeting all of these goals will be outlined in the remainder of this paper.

Transform Summary. The layout of our long-range transform is shown in the flow chart of Figure 3. The main characteristics of the transform are discussed separately in the next paragraphs, but a brief introduction is needed to get a picture of the data flow in the transform. The transform starts off with a quick check of whether or not the target is

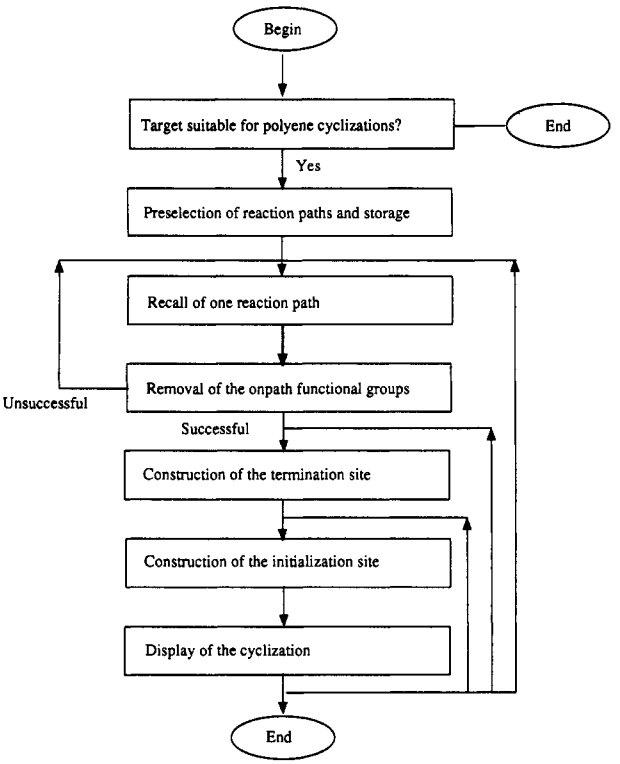

Figure 3. Flow chart of the polyene cyclizations transform.

suitable to be synthesized by a polyene cyclization. If the target passes this first test, the skeleton (either carbocyclic or heterocyclic) of the molecule is examined for the best places to disconnect skeleton bonds. The result is a list of so-called reaction paths which range from the termination site to the initiation site, including all the bonds in between which participate in the reaction. These selected reaction paths are stored and subsequently recalled one by one. The paths are checked for interfering groups, and each interfering group has to be exchanged for an acceptable one, using subgoal transforms. If there is no chance to use the group as a part of the initiation or termination site, then the group will be removed in this stage of the analysis. Subsequently, an attempt is made to introduce the first of the main termination sites using subgoal transforms, followed by a systematic search for a suitable initiation site. If all initiation sites have been considered in combination with a certain termination site, then the next termination site will be applied (the entire process can be seen in Figure 3). Each successful result from this analysis will be displayed to the user immediately.

DEVELOPMENT OF A NEW LONG-RANGE TRANSFORM: DETAILS

Evaluation of the Target Molecule. It is by no means certain that the user-specified molecule, which is presented to the system as a target, can be the product of a polyene cyclization, even if this method is selected by the user. So, a filter is needed to eliminate all target molecules which are definitely unsuitable for polyene cyclizations (e.g., acyclic compounds). In most other long-range transforms the standard method for coding such a filter is the description of a required ring substructure in a specially developed syntax[23] (similar to the retron description used in normal LHASA transforms). This method suffices for reactions producing a predefined substructure, but it will fail in the case of the cationic polyene cyclizations due to the variety of ring structures among the possible target molecules. As a consequence, the filter, or alternatively stated, the keying information of this long-range

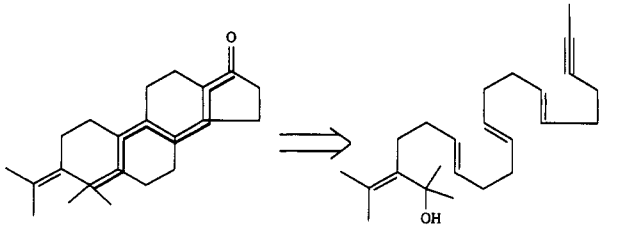

Figure 4. Reaction path of a polyene cyclization (marked in bold).

transform, had to be programmed separately, and we decided to implement this code in the LHASA program (FORTRAN code) and not in the knowledge base (CHMTRN) in order to use the power of the LHASA perception module. Each target submitted to LHASA is perceived by scanning it on a variety of chemically important properties, as rings, ring sizes, heteroatoms, bridges, etc. The information collected in this survey is stored bitwise. An 8-bytes computer word (called set) is reserved for each property with the *n*th bit turned on if atom *n* has the property of interest. Most properties require two sets to record unambiguously all the acquired information: one set to save the atoms having that property and one set to save the bond information. An obvious consequence of this type of storage is the current upper limit of 64 to both the number of atoms and the number of bonds in a LHASA molecule. This perception information is the basis for deciding on the applicability of the polyene cyclization transform.

The observation that polyene cyclizations generally form fused ring systems is an important clue in selecting a filter for target molecules for polyene cyclizations. In practice, the size of the rings involved in the cyclization varies from 5 to 7; hence, we used the presence of a fusion between two five-to seven-membered rings as the criterion for identifying suitable targets for a polyene cyclization. This criterion was implemented by using the sets with information about ring sizes and fusion bonds stored in the perception phase of the LHASA analysis. After unsuitable targets have been eliminated, the next step in the analysis focuses on practically useful cyclizations.

Selection and Storage of Reaction Paths. The vast majority of theoretically possible cyclizations can be rejected because of the inaccessibility of the evolving precursors, as most of them are at best as difficult to synthesize as the target molecule. For program efficiency the few useful cyclizations should be identified as early in the analysis as possible, and therefore our transform initially virtually ignores the substitution pattern of the target and focuses entirely on the skeleton (either carbon or heterocyclic). A wide range of skeletons can be generated by polyene cyclizations, mainly depending on the type of initiating and terminating group used. A survey of literature precedents[24-28] of polyene cyclizations which were successfully applied to the synthesis of natural products shows that they produce an alternating chain of fusion and nonfusion bonds in the vast majority of the literature examples. The fusion bonds stem from the double bonds in the polyene, while the bonds in between are newly formed (see Figure 4). This observation gave us a clue for identifying the reaction path for useful cyclizations. The bonds which are retrosynthetically broken in a polyene cyclization constitute an alternating chain with bonds to be doubled, and this feature of polyene cyclizations, translated into CHMTRN code, was used to find the reaction path.

Because some structural features such as bridgeheads, fusions spanning small (three- or four-membered) rings, or large (eight-membered or more) rings, obstruct the polyene cyclization, all fusions in the target molecule are examined

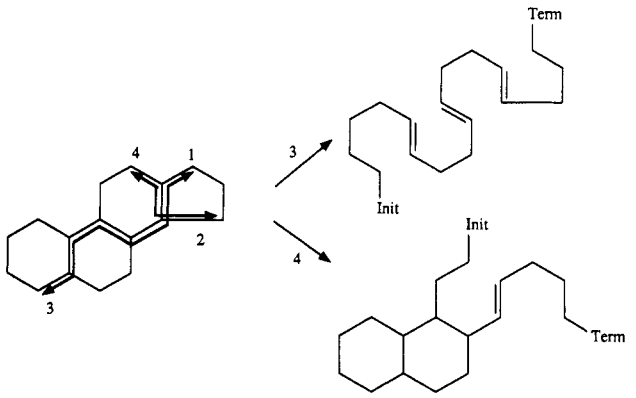

Figure 5. Bond selected as termination site initially, giving rise to four potential reaction paths (shown as 1–4).

to reject the ones containing one of these unfavorable features. The remaining fusions are stored in a CHMTRN set, an efficient way to store a number of atoms or bonds in the molecule as an entity. These sets are similar to the sets created by the FORTRAN code in the perception phase. Only fusions included in the set of potential reaction path fusions will be used to construct reaction paths. Such a reaction path needs to be "grown" from one of its ends, either the initiation site or the termination site. Since termination sites frequently can be identified by the ring system (e.g., five-membered rings,[3,7] aromatic rings[29,30]), the choice has been made to use the termination site for this purpose.

All fusions stored in the set of promising ones are judged on their merits as a termination site. The first consideration for finding the termination site is the ring size. Endocyclic cyclization at an olefin is a process which is disfavored according to the Baldwin rules.[31] Therefore, the formation of five-membered rings requires the use of dedicated methods[7] and will usually result from the termination of the polyene cyclization. Hence, fusions on a five-membered ring are obvious candidates for the termination site. A second consideration used to select a potential termination site is the length of the reaction path. A longer reaction path corresponds to a polyene cyclization that accomplishes a more significant simplification, although the yield is decreasing with increasing path length.[32] Since six-membered rings,[3-7] and to some extent seven-membered rings,[24,33] are perfectly suitable in acting as the central ring in a polyene cyclization, these rings are omitted from the set of termination sites if they contain a 1,3-difusion (and both fusions are included in the set of suitable ones).

Each potential termination site, tracked down as described above, is the starting point of a reaction path. With use of the knowledge about the topology of the target obtained during the perception phase, a straightforward CHMTRN procedure was written to obtain the reaction path, although some targets containing doubly fused atoms (e.g., rings with 1,2-difusions or propellanes) required some extra attention. One more evaluation had to be carried out on the reaction path, since the actual termination site is not located on the fusion itself, but on an adjacent atom. So, each fusion selected as potential termination site actually consists of four termination sites. Each of these sites corresponds to a reaction path growing in a different direction (shown as 1–4 in Figure 5). Fortunately, the previous selection procedure of the termination sites resulted not only in a fusion bond but also in the ring which has to contain the actual termination site. So, two of the four alternative paths can be ignored immediately (the ones indicated by 1 and 2 in Figure 5). The other two may be equally reasonable, but usually one of them is by far more strategic than the other, as illustrated by Figure 5 for the case

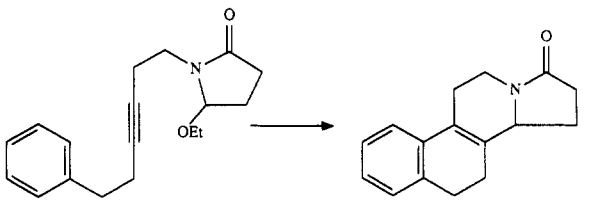

Figure 6. Literature example of an olefinic path bond.

of a steroid target. The precursor resulting from 4 is not significantly simpler than the target molecule itself, but the precursor resulting from reaction path 3 is a big step forward to simpler starting materials. Each reaction path resulting from this analysis is stored in CHMTRN sets to be converted to real chemistry in the remainder of the transform.

Usually, the molecule skeleton analysis results in two reaction paths, which are the reverse of each other. A typical example is the steroid skeleton of Figure 5. One reaction path has the termination site at the A-ring, while the other one has the termination site at the D-ring. This result corresponds with the literature which shows examples of cyclizations of both types.[28,30,34,35]

Recall of One Reaction Path. The analysis described thus far has been focused entirely on the topology of the target (ignoring the details of the reactions needed) to find the most strategic disconnections in the target. At this stage, it is time to restore separately each of the strategic paths stored in the first part of the transform and translate them into actual chemical reactions. This translation cycle will usually demand some subgoal steps to set up the termination and initiation. For each reaction path, a perception of that path is performed and the resulting information is stored in several sets. The main result of this perception is the exact constitution of the path: the number of bonds on the path (path length), the stereochemistry of the fusions, and the location of the fusion bonds on the path.

Removal of On-Path Functional Groups. The reaction path must be free of functional groups except for the ones involved in the termination and initiation, since functional groups will usually interfere with the reaction. Only two exceptions to this rule are known. The first one is a fluoride,[36] which has a surprisingly positive effect on the reaction instead of the inhibition that might be expected; the second one is an olefin, which is shown to be the product of some acetylenic polyenes[25,37] (see Figure 6 for an example[37]). Usually, an acetylene will terminate the reaction, resulting in a five-membered ring,[7] but in some cases the reaction can be directed toward six-membered ring formation.[25,37] If the center of the reaction path contains functional groups other than olefin or fluoride, some kind of a functional group removal will be necessary. However, a full (retrosynthetic) removal is usually not feasible, as synthetically it implies a selective introduction of a functional group on a certain position in the molecule. So, the only way to remove an interfering functional group is an exchange for either an olefin or a fluoride. We have chosen the olefin for this purpose, as the olefin group is a better source for creating functional groups other than the fluorine group. Subsequently, the entire path is checked on functional groups, and only functional groups at both ends of the path are left unchanged, as they are expected to be interchangeable for a suitable initiation or termination site.

Initiation and Termination. A successful polyene cyclization is virtually impossible without powerful initializing and terminating groups.[3-7] The long-range transform described in this paper meets this requirement by demanding accurately defined substructures at the termination and initializing sites.

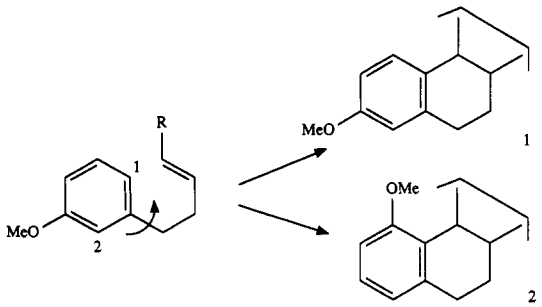

Figure 7. Termination site of an aromatic group terminated polyene cyclization. The freedom of rotation around the marked bond leads to reaction at both positions 1 and 2.

Usually, these substructures cannot be found in the target molecule itself, but subgoal transforms offer a powerful tool for converting the target molecule into a compound containing the desired substructures. The management of the cascade of subgoals necessary for converting the target into a structure having a certain initializing or terminating site has been implemented in separate subroutines for each type of these sites. We chose this approach to guarantee a well-ordered structure of the transform. Furthermore, it opens the possibility of applying the carbon skeleton analysis developed for the cationic polyene cyclizations to other types of polyene cyclizations (notably radical[38] and palladium catalyzed[39] cyclizations). The development of long-range transforms for these types of polyene cyclizations is from now on mainly limited to the development of a new set of subroutines to describe the appropriate methods for the initiation and termination.

Each subroutine directing the transformations to a certain initiation or termination site does not include a complete description of all the subgoal steps involved. Such an approach would be highly inefficient and probably not even feasible as the variation of substructures found in target molecules is virtually infinite. Instead, the subroutines mainly direct the existing subgoal system of LHASA to the desired result. Exact reaction descriptions have only been included for those methods which are not covered by the standard subgoal transforms, which usually concerned specific methods developed to convert the product of a polyene cyclization to popular natural products, such as progesterone. Methods which are missing from the existing subgoal system, but for which a broader use could be foreseen, have been written as new subgoal transforms.

The details specific to the implementation of each termination or initiation site will be discussed in the next paragraphs.

Termination Sites: Aromatic Rings. Termination by aromatic substructures has a widespread use[25,28,29] with satisfactory results, provided that the aromatic ring is not too deactivated. A regioselectivity problem has to be considered, as the polyene system has freedom of rotation around the bond indicated in Figure 7 to react on both ortho sites of the aromatic ring, resulting in different products in the case of asymmetric rings.[30] Both the factors controlling the degree of deactivation and the factors determining the regioselectivity are similar to the factors which control analogous problems with electrophilic aromatic substitutions. These substitutions were already treated in detailed CHMTRN subroutines performing Hammett calculations, but a direct use of these subroutines in the new long-range transform was hindered by the fact that intramolecular reactions were originally not considered in the implementation of the Hammett equations. Intramolecular reactions require a different approach since the meta and para attack need not to be considered, as the resulting products, *m*- and *p*-cyclophane, respectively, are usually too strained. This limitation of the implementation of the electrophilic aromatic substitution reactions in LHASA was removed, allowing its use in the polyene cyclization transform.

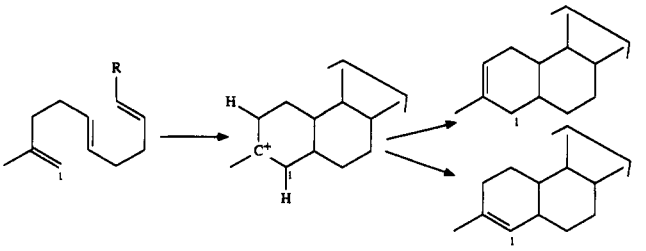

Figure 8. Termination by proton abstraction.

Termination Sites: Olefins. The final intermediate in a fully stepwise description of polyene cyclizations will consist of a fully closed ring system containing a cation on the previously closed ring (see also Figure 8). A popular method for trapping this cation is the abstraction of a proton from a neighboring carbon, leading to a double bond between these two centers.[3,7] This olefin is the characteristic product of this type of termination and should therefore be present in the target molecule. Most target molecules, however, lack the essential olefin, so subgoal transforms are needed to introduce it. An important problem is where to introduce the olefin. Several positions on the terminating ring have to be considered, both endocyclic[36,40] and exocyclic.[27,32,41] All atoms α to the last atom of the reaction path (atom 1 in Figure 8) are possible hosts of the intermediate cation. Theoretically each of the α atoms of the intermediate cation carrying a proton (including the ipso position) can be the source for the deprotonation. The (retro)reaction sequences to introduce these olefins differ with the position of the olefin in the molecule, but it would be annoying and confusing to show all these variations to the user. Therefore, the long-range transform selects the most promising variations. For this purpose, a system has been devised which considers the synthetic accessibility of alternative precursors and compares the probability of different regioisomers. A cation stability rating is assigned to each bond site which is a potential place for the olefinic bond in the precursor. The rating is computed based on the carbon substitution (primary vs secondary, etc.) and the olefin stabilization which can be reached on this bond (allylic, benzylic, etc). Usually, it is not sufficient to compare only the stability of the possible olefins, as the olefin must be introduced using functional group manipulations (via subgoal transforms). The ease of this process differs considerably from one site to another. This factor is given the higher priority in the transform, as the olefin will be introduced at the spot containing some functionality already (if possible). If this site is not the same as the site with the highest calculated cation stability rating, a trialkylsilyl group can be used to direct the regioselectivity of the reaction.[32,41,42] Dependent on the desired product, a vinyl silane or an allyl silane will be used. Sometimes, more than one of the possible precursors will have a high tactical value. In that case both solutions will be displayed using trialkylsilyl groups in the case of the product that would otherwise not be expected to be the main product.

Termination Sites: Trapping by the Solvent. Termination of a polyene cyclization using olefins produces a fully ring-closed intermediate. This intermediate can be subjected to deprotonation as described in the previous paragraph, or it can react with protic agents like TFA.[43,44] The resulting trifluoroacetate can be hydrolyzed to an alcohol. This method is somewhat more limited than the previous one and will usually

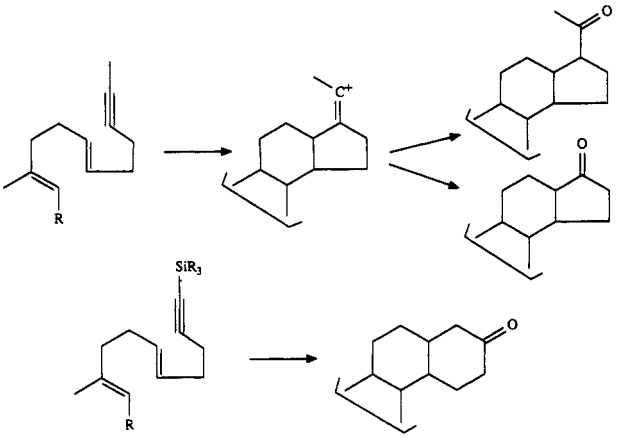

Figure 9. Three different termination methods using acetylenes.

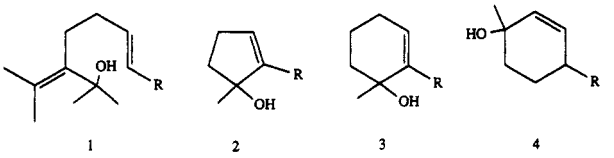

Figure 10. Four implemented allylic alcohol initiating groups.

be accompanied by olefin formation.[43] For that reason, the long-range transform only uses this termination type in the case of functionalities which can easily be converted to the acetate moiety.

Termination Sites: Acetylenes. Usually a cationic cyclization will lead to six-membered rings only, but cyclization with acetylenes has a high preference for cyclization to five-membered rings. The (hypothetical) intermediate is a vinyl cation (see Figure 9) which can produce an exocyclic ketone on work-up with a protic solvent,[26] as the resultant enol ether intermediate can be hydrolyzed to a ketone. On the other hand, an oxidative work-up can be used to produce an endocyclic ketone.[45] A further improvement to the flexibility of the acetylene termination method is obtained by the use of the trialkylsilyl group, as alkynylsilyl groups produce six-membered rings containing an endocyclic ketone[34] (second reaction in Figure 9). All these variations have been implemented in the long-range transform.

Initiation Sites: Allylic Alcohols. Allylic alcohols are the most popular among the initiation groups.[3,7] Although they are extremely useful in many cases, the differences between powerful and nonreacting allylic alcohols are very subtle. Small structural differences determine the success or failure of the polyene cyclization. The allylic alcohols used successfully in polyene cyclizations can be divided into four classes,[3] which are shown in Figure 10.

The new long-range transform will use its subgoal power to reach one of these classes of allylic alcohols. It is important to note that the first three types of allylic alcohols suffer from a regioselectivity problem as the allylic cation formed initially is symmetrical, allowing reaction at both ends of the allylic system. The long-range transform will notice this complication and display a warning to alert the user.

Initiation Sites: Other Initiating Groups. Although the allylic alcohols are by far the most frequently used initiation groups, a few other groups have gained some degree of popularity. Four of them (shown in Figure 11) have been considered of sufficient importance to implement them in the long-range transform. The oldest one is the epoxide group developed mostly by van Tamelen[46,47] in the 1960s and the 1970s. The idea of this initiation has been derived from the biosynthesis of Lanosterol from Squalene, already shown in

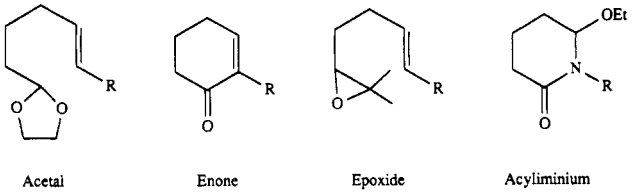

Figure 11. Other initiating groups.

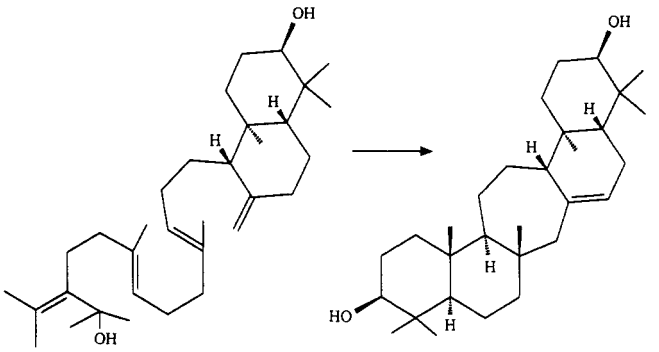

Figure 12. Synthesis of serratenediol using polyene cyclization methodology.

Figure 1. Some examples in the syntheses of natural products have been reported,[35] although the yields generally remain rather low.

Apart from the work on the allylic alcohol groups, the Johnson group also developed the acetal group[40] as an initiating site. The major limitation of this group, is the same as for the previous one, the restriction of six-membered rings. The work of Harding[48] led to the introduction of the enone as an initiating group. The use of this group is limited to six-membered rings for unclear reasons. Still, we considered the group of sufficient relevance to justify implementation in the long-range transform. The fourth group is the acyl–iminium group (Figure 11) developed by Speckamp and co-workers,[37,49] which is an elegant way to build in a nitrogen atom in the polyene skeleton. It is the only method which has attracted some attention to incorporate heteroatoms in the reaction path.

APPLICATIONS OF THE NEW TRANSFORM

We have analyzed two typical examples of polycyclic molecules to show that the transform is able to reproduce literature syntheses and can generate reaction routes to polycyclic compounds including steroids, for which the biomimetic approach was originally designed. The first example is the synthesis of serratenediol, a pentacyclic triterpenoid. The second example shows the synthesis of a precursor of progesterone.

Example 1. Serratenediol (**1**), a pentacyclic triterpenoid, was one of the first natural products synthesized using polyene cyclization methodology. The original synthesis, published in 1974,[27] is shown in Figure 12. We have analyzed this molecule in LHASA using our new cationic polyene cyclization transform, expecting that the transform would be able to produce some alternative sequences, one of which could be the literature synthesis. In fact, LHASA came up with three sequences (Figure 13), of which the first one (leading to the polyenic precursor **4**) is very similar to the original synthesis. This three step sequence shows the smooth interaction between the long-range transform and the LHASA subgoal transforms. The initiation is set up by the subroutine guiding allylic alcohol initiations, which calls subgoal transforms from the core knowledge base to perform functional group manipulations when possible. The first step is a good example, showing the

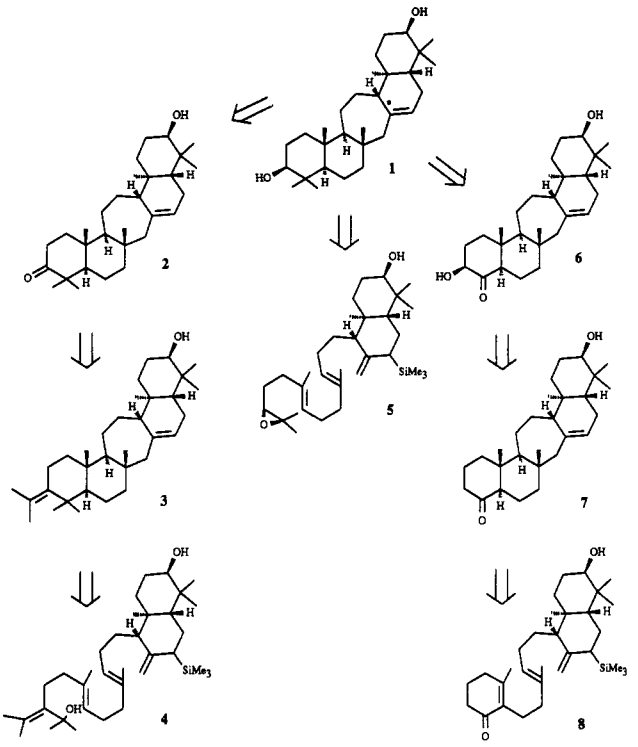

Figure 13. LHASA analysis of serratenediol.

ketone reduction needed to convert alcohol **1** to ketone **2**. Reactions too specific to be included in the set of subgoal transforms have been included in the subroutine itself. The second step, the introduction of the isopropylidene group, is an example of such an "in-table" subgoal.

The termination does not require any change in the target, as the olefin termination can be applied directly, but the regioselectivity requires special attention. The subroutine dealing with olefin terminations notices another likely proton abstraction which is the one leading to an olefin on the fusion marked with an asterisk in structure **1**. For that reason, the right regioselectivity is enforced using a trialkylsilyl group. This concern about the regioselectivity is not supported by the original paper,[27] as the regioisomer is not reported as a side product while the silicon stabilization is not used in practice. However, the yield is rather low (20%) and no detailed analysis of the side products is given, but in some similar cases a mixture of products is observed,[3] which could have been avoided using the trialkylsilyl group.[32] The second and third sequence show alternative initiation sites; the second one uses an epoxide and the third one an enone. The epoxide can be used directly without any subgoal steps; the enone requires some more work. The first step is a transform (dimethylation of a ketone) which actually consists of several steps: A Wittig reaction followed by a Simmons–Smith reaction to set up a cyclopropane, which can be reduced to the dimethyl. The precursor of this transform is ketone **6**, which contains a potentially interfering α-hydroxy group. The subroutine governing the enone initiation notices this possible problem and calls a functional group removal transform, an oxidative hydroxylation, to (retrosynthetically) remove the alcohol. No alternative termination sites have been found since the seven-membered ring obstructs all other types of terminations. The path selection process resulted in another reaction path (the one with the termination on the A-ring and the initiation on the C-ring), but this reaction path is not successful, as none of the initiating groups can operate on a seven-membered ring.

Example 2. The biomimetic approach has been designed originally as a synthetic method to obtain steroids and steroid-like compounds. Figure 14 shows our analysis of such a compound (structure **9**), a precursor of norprogesterone. This analysis shows the effect of the reaction path selection, as implemented in our transform. The transform selects two strategic paths in **9**, one path having the initiation at the A-ring and termination at the D-ring and the other one having them interchanged. LHASA finds sequences for both reaction paths when structure **9** is analyzed, and both reaction paths have been realized in syntheses of steroid-like molecules.[28,30,34,35] The first three sequences (the ones with respectively precursors **11**, **14**, and **19** as the precursor of the first step) have the initiating site located at the D-ring, and they differ only in the termination site. All three sequences need one or two subgoal steps to introduce a suitable functionality for termination, which is accomplished by core knowledge base subgoal transforms directed by the new long-range transform. The initiation requires more work but is identical for the three sequences and is for that reason only drawn for the second sequence. The exocyclic ketone is removed using a methoxy–ylide Wittig transform. The methyl group is blocking the allylic alcohol initiating site and is removed using a dedicated method, the BF_3-catalyzed rearrangement,[30] which is directly described in the allylic alcohol subroutine. For the subsequent epoxidation, the transform from the core knowledge base can be used, which finally paves the road for the actual polyene cyclization.

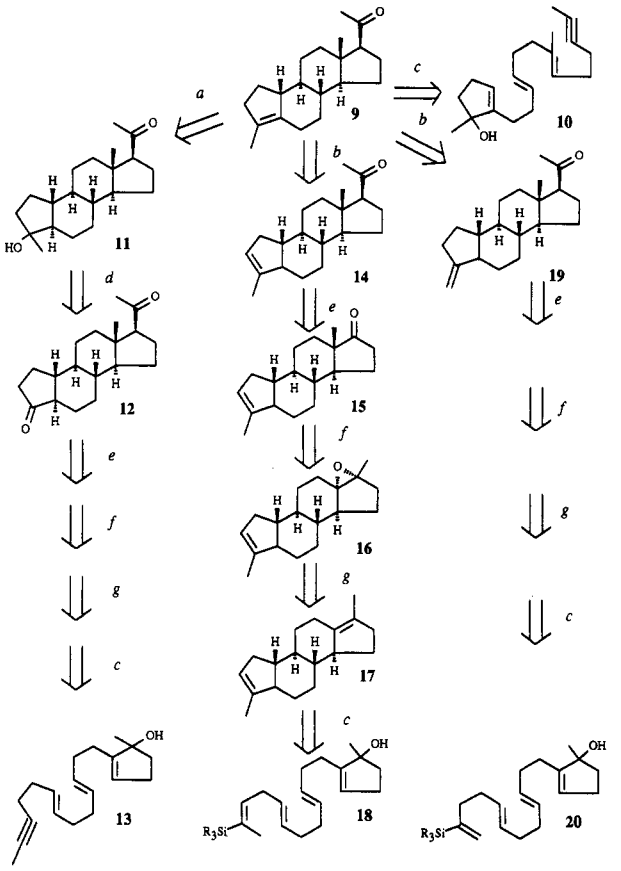

Figure 14. Analysis of a precursor of norprogesterone: (a) dehydration; (b) olefin isomerization; (c) polyene cyclization; (d) Grignard; (e) Wittig via methoxy–ylide; (f) BF_3-catalyzed rearrangement; (g) epoxidation.

Only after all solutions have been found for one reaction path, the transform will go over to the next reaction path. This means that the fourth and shortest solution (leading to precursor **10**) is found as the last one. This route has been derived from the other reaction path with the initiating group

at the A-ring. This reaction is identical to the original one in the paper of the Johnson group.[50]

CONCLUSIONS

Polycyclic compounds like steroids and terpenes are exactly the types of molecules for which LHASA was originally designed. Nevertheless, our LHASA analysis of steroid molecules showed a serious shortcoming in the program's approach to the synthesis planning of these compounds. It was effectively impossible to create polyene cyclizations with the system, due to the absence of a strategic analysis for this type of reaction. This diagnosis of the problem led us to the development of a long-range transform which determines, on the basis of the carbon skeleton, which polyene cyclizations are expected to realise the largest simplification of the molecule. Only after the assignment of these most strategic cyclizations, the transform starts to look for actual chemical reactions which can perform the selected polyene cyclization. This translation to real chemical reactions required the introduction of a great deal of specific knowledge from the literature, to ensure that the transform focuses only on chemically realistic cyclizations and does not omit important cyclization routes. The examples show that the transform not only is able to reproduce literature syntheses but also can produce alternative synthesis routes.

The design of the transform easily allows the input of knowledge about newly developed initiating or terminating groups when future publications give rise to such an action. Furthermore, our approach can be applied to other types of chain reactions when these get a widespread use in the synthesis of natural products. Most other than cationic polyene cyclizations chain reactions currently known (notably radical reactions) have a preference for the formation of somewhat different ring systems. Hence, the detection of the strategic paths on the basis of an alternating chain of fusion and nonfusion ring bonds might not be sufficient for these cases. However, the general approach outlined in this paper will still be applicable to another type of chain reaction and can easily be implemented as a long-range transform in the LHASA system.

ACKNOWLEDGMENT

We are grateful to M. A. Ott (CAOS/CAMM Center) for his valuable support. Use of the services and facilities of the Dutch CAOS/CAMM Center, under grant Nos. SON 326-052 and STW NCH99.1751, is gratefully acknowledged. J.-W.B. thanks Organon International BV for financial support.

REFERENCES AND NOTES

(1) Cornforth, J. W. Olefin-Alkylierung in der Biosynthese. *Angew. Chem.* **1968**, *80*, 977–1008; *Angew. Chem., Int. Ed. Engl.* **1968**, *7*, 903.
(2) McCarry, B. E.; Markezich, R. L.; Johnson, W. S. Direct Formation of the Steroid Nucleus by a Nonenzymic Biogenetic-Like Cyclization. Cyclization and Proof of Structure and Configuration of Products. *J. Am. Chem. Soc.* **1973**, *95*, 4416–4417.
(3) Johnson, W. S. Biomimetic Polyene Cyclizations. A Review. *Bioorg. Chem.* **1976**, *5*, 51–98.
(4) van Tamelen, E. E. Bioorganic Chemistry: Total Synthesis of Tetra- and Pentacyclic Triterpenoids. *Acc. Chem. Res.* **1975**, *8*, 152–158.
(5) Sutherland, J. K. The Initiation of Cyclization Using 3-Methylcyclohex-2-enone Derivatives. *Chem. Soc. Rev.* **1980**, *9*, 265–280.
(6) Bartlett, P. A. Olefin Cyclization Processes That Form Carbon-Carbon Bonds. *Asymmetric Synth.* **1984**, *3*, 341–411.
(7) Johnson, W. S. Biomimetische Cyclisierungen von Polyenen. *Angew. Chem.* **1976**, *88*, 33–66; *Angew. Chem., Int. Ed. Engl.* **1976**, *15*, 9.
(8) Dewar, M. J. S.; Reynolds, C. H. π-Complexes as Intermediates in Reactions. Biomimetic Cyclization. *J. Am. Chem. Soc.* **1984**, *106*, 1744–1750.
(9) Ladika, M.; Bregovec, I.; Sunko, D. E. On the Nonconcertedness of Allylic Cation Promoted π-Cyclization Reactions. *J. Am. Chem. Soc.* **1981**, *103*, 1285–1286.
(10) Nishizawa, M.; Takenaka, H.; Hayashi, Y. Experimental Evidence of the Stepwise Mechanism of a Biomimetic Olefin Cyclization: Trapping of Cationic Intermediates. *J. Am. Chem. Soc.* **1985**, *107*, 522–523.
(11) Saito, A.; Matsushita, H.; Kaneko, H. Cyclization Mechanism of Trans-β-Monocyclohomofarnesic Acid and Related Compounds. *Chem. Lett.* **1983**, 729–732.
(12) Corey, E. J.; Long, A. K.; Rubenstein, S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408–418.
(13) Ott, M. A.; Noordik, J. H. Computer tools for reaction retrieval and synthesis planning in organic chemistry. A brief review of their history, methods, and programs. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 239–246.
(14) Zefirov, N. S.; Gordeeva, E. V. Computer-assisted Synthesis. *Russ. Chem. Rev.* **1987**, *56*, 1002–1014.
(15) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses. *Science* **1969**, *166*, 178–192.
(16) Corey, E. J.; Long, A. K.; Lotto, G. I.; Rubenstein, S. D. Computer-assisted synthetic analysis. Quantitative assessment of transform utilities. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304–309.
(17) Corey, E. J. Die Logik der chemischen Synthese: Vielstufige Synthesen komplexer carbogener Moleküle (Nobel-Vortrag). *Angew. Chem.* **1991**, *103*, 469–479; *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 455.
(18) Corey, E. J. Retrosynthetic Thinking—Essentials and Examples. *Chem. Soc. Rev.* **1988**, *17*, 111–133.
(19) Corey, E. J.; Cheng, X. M. *The Logic of Chemical Synthesis*; Wiley: New York, 1989.
(20) Corey, E. J.; Jorgensen, W. L. Computer-Assisted Synthetic Analysis. Generation of Synthetic Sequences Involving Sequential Functional Group Interchanges. *J. Am. Chem. Soc.* **1976**, *98*, 203–209.
(21) Corey, E. J.; Johnson, A. P.; Long, A. K. Computer-Assisted Synthetic Analysis. Techniques for Efficient Long-Range Retrosynthetic Searches Applied to the Robinson Annulation Process. *J. Org. Chem.* **1980**, *45*, 2051–2057.
(22) Corey, E. J.; Long, A. K.; Mulzer, J.; Orf, H. W.; Johnson, A. P.; Hewett, A. P. W. Computer-Assisted Synthetic Analysis. Long-Range Search Procedures for Antithetic Simplification of Complex Targets by Application of the Halolactonization Transform. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 221–230.
(23) Corey, E. J.; Howe, W. J.; Pensak, D. A. Computer-Assisted Synthetic Analysis. Methods for Machine Generation of Synthetic Intermediates Involving Multistep Look-Ahead. *J. Am. Chem. Soc.* **1974**, *96*, 7724–7737.
(24) Groen, M. B.; Zeelen, F. J. Biomimetic total synthesis of steroids VIII. Stereospecific synthesis of B-homo-19-norsteroids. *Recl. Trav. Chim. Pays-Bas* **1984**, *103*, 169–173.
(25) Bartlett, W. R.; Johnson, W. S.; Plummer, M. S.; Small, V. R. Jr. Cationic Cyclization of a Substrate Having an Internal Acetylenic Bond. Synthesis of Euphol and Tirucallol. *J. Org. Chem.* **1990**, *55*, 2215–2224.
(26) Gravestock, M. B.; Johnson, W. S.; McCarry, B. E.; Parry, R. J.; Ratcliffe, B. E. Acetylenic Bond Participation in Biomimetic Polyene Cyclizations. Model Studies Directed toward the Synthesis of 20-Keto Steroids. Synthesis of *dl*-Progesterone and *dl*-Δ^4-Androstene-3,17-dione. *J. Am. Chem. Soc.* **1978**, *100*, 4274–4282.
(27) Prestwich, G. D.; Labovitz, J. N. Application of Nonenzymic Biogenetic-Like Olefinic Cyclizations to the Total Synthesis of *dl*-Serratenediol. *J. Am. Chem. Soc.* **1974**, *96*, 7103–7105.
(28) Bartlett, P. A.; Johnson, W. S. A Stereospecific Total Synthesis of Estrone *via* a Cationic Olefinic Cyclization. *J. Am. Chem. Soc.* **1973**, *95*, 7501–7502.
(29) Macco, A. A.; de Brouwer, R. J.; Buck, H. M. Asymmetric Induction in the Synthesis of Thiophene-Containing Steroidlike Molecules via Olefinic Cyclization. Precoiling as Model Description for the Stereochemical Course of the Reaction. *J. Org. Chem.* **1977**, *42*, 3196–3199.
(30) Groen, M. B.; Zeelen, F. J. Biomimetic Total Synthesis of 14α-Methyl-19-Norsteroids. Stereoselective Epoxidations with $Mo(CO)_6$/t-BuOOH. *Tetrahedron Lett.* **1982**, *23*, 3611–3614.
(31) Baldwin, J. E.; Cutting, J.; Dupont, W.; Kruse, L.; Silberman, L.; Thomas, R. C. 5-*Endo*-*Trigonal* Reactions: a Disfavoured Ring Closure. *J. Chem. Soc., Chem. Commun.* **1976**, 736–738.
(32) Johnson, W. S.; Chen, Y.-Q.; Kellogg, M. S. Termination of Biomimetic Cyclizations by the Allylsilane Function. Formation of the Steroid Nucleus in One Step from an Acyclic Polyenic Chain. *J. Am. Chem. Soc.* **1983**, *105*, 6653–6656.
(33) Groen, M. B.; Hindriksen, B.; Zeelen, F. J. Biomimetic Total Synthesis of Steroids, IX. Stereospecific Synthesis of C-homo-19-norsteroids. *Recl. Trav. Chim. Pays-Bas* **1985**, *104*, 59–62.
(34) Johnson, W. S.; Yarnell, T. M.; Myers, R. F.; Morton, D. R.; Boots, S. G. Biomimetic Polyene Cyclizations. Participation of the (Trimethylsilyl)acetylenic Group and the Total Synthesis of the *D*-Homosteroid System. *J. Org. Chem.* **1980**, *45*, 1254–1259.
(35) van Tamelen, E. E.; Hwu, J. R. Direct Total Synthesis of Traditional Sterols by Tricyclization of Polyunsaturated Cyclohexene Oxides. *J. Am. Chem. Soc.* **1983**, *105*, 2490–2491.
(36) Johnson, W. S.; Chenera, B.; Tham, F. S.; Kullnig, R. K. The Fluorine Atom as a Cation-Stabilizing Auxiliary in Biomimetic Polyene Cyclizations. 1. Background and Exploratory Experiments. *J. Am. Chem. Soc.* **1993**, *115*, 493–497.
(37) Dijkink, J.; Speckamp, W. N. Biomimetic Heterocyclisation of Aryl Olefins One-Step Formation of Two Carbon-Carbon Bonds. *Tetrahedron* **1978**, *34*, 173–178.

(38) Curran, D. P. The Design and Application of Free Radical Chain Reactions in Organic Synthesis. Part I. *Synthesis* **1988**, 417–439.
(39) Trost, B. M.; Shi, Y. A Pd-Catalyzed Zipper Reaction. *J. Am. Chem. Soc.* **1991**, *113*, 701–703.
(40) Johnson, W. S.; Wiedhaup, K.; Brady, S. F.; Olson, G. L. Nonenzymic, Biogenetic-Like Cyclization of a Tetraenic Acetal to Produce the *D*-Homosteroid Nucleus. *J. Am. Chem. Soc.* **1974**, *96*, 3979–3981.
(41) Johnson, W. S.; Newton, C.; Lindell, S. D. The Carboalkoxyallylsilane Terminator for Biomimetic Polyene Cyclizations. A route to 21-Hydroxyprogesterone Types. *Tetrahedron Lett.* **1986**, *27*, 6027–6030.
(42) Blumenkopf, T. A.; Overman, L. E. Vinylsilane- and Alkynylsilane-Terminated Cyclization Reactions. *Chem. Rev.* **1986**, *86*, 857–873.
(43) Johnson, W. S.; Dawson, M. I.; Ratcliffe, B. E. Biomimetic Polyene Cyclizations. Synthesis and Cyclization of 1,3-Dimethyl-2-(3-methyl-*trans*-3,7-octadienyl)cyclohex-2-en-1-ol. *J. Org. Chem.* **1977**, *42*, 153–158.
(44) Harding, K. E.; Leopold, E. J.; Hudrlik, A. M.; Johnson, W. S. Cyclization of 4-(*trans*-3,7-Octadienyl)-3-methyl-2-cyclohexen-1-ol and 4-(*trans*,*trans*-7-methyl-3,7,11-dodecatrienyl)-3-methyl-2-cyclohexen-1-ol *J. Am. Chem. Soc.* **1974**, *96*, 2540–2549.
(45) Johnson, W. S.; Ward, C. E.; Boots, S. G.; Gravestock, M. B.; Markezich, R. L.; McCarry, B. E.; Okorie, D. A.; Parry, R. J. Formation of Vinyl Halides from Vinyl Cations Generated by Acetylenic Participation in Biomimetic Polyene Cyclizations. *J. Am. Chem. Soc.* **1981**, *103*, 88–98.
(46) van Tamelen, E. E. Bioorganic Chemistry: Sterols and Acyclic Terpene Terminal Epoxides. *Acc. Chem. Res.* **1968**, *1*, 111–120.
(47) van Tamelen, E. E.; James, D. R. Overall Mechanism of Terpenoid Terminal Epoxide Polycyclizations. *J. Am. Chem. Soc.* **1977**, *99*, 950–952.
(48) Cooper, J. L.; Harding, K. E. Cationic Cyclization of α-β-Unsaturated Ketones. A Facile Synthesis of 9-Methyldecalin-2-ol-5-one, an Intermediate for Synthesis of Eudesmane Sesquiterpenoids. *Tetrahedron Lett.* **1977**, *3321*–3324.
(49) Speckamp, W. N.; Hiemstra, H. Intramolecular Reactions of N-Acyliminium Intermediates. *Tetrahedron* **1985**, *41*, 4367–4416.
(50) Johnson, W. S.; Huffman, W. F.; Boots, S. G. Biomimetic polyene cyclizations. Synthesis of racemic 19-norprogesterone. *Recl. Trav. Chim. Pays-Bas* **1979**, *98*, 125–126.