Bull Soc Chim Fr (1997) 134, 669-676

Computer-oriented representation of stereoselective synthetic methods

Andreas Dietz[1]*, Jacques Coste[2], Claude Laurenço[2]

[1] *Laboratoire d'Informatique, de Robotique et de Microélectronique, UMR 5506 du CNRS et de l'Université Montpellier II, GDR 1093 du CNRS, 161, rue Ada, 34392 Montpellier cedex 5;*
[2] *Centre CNRS-INSERM de Pharmacologie-Endocrinologie, UPR 9023 et GDR 1093 du CNRS, 141, rue de la Cardonille, 34094 Montpellier cedex 5, France*

(Received 21 January 1997; accepted 7 July 1997)

Summary — Computer programs which are designed to assist in the planning of organic syntheses can suggest pertinent solutions only if they use an adequate computer-oriented representation of synthetic methods. In order to generate synthetic precursors from a given target, these programs use retrosynthetic representations of synthetic methods, so-called transforms. Although stereochemistry is an essential factor of synthetic complexity, up to now only few efforts have been made to represent the stereochemical aspect of transforms. The known approaches exclusively rely on basic knowledge about the stereochemical course of reactions, and every transform is described in terms of a few elementary stereochemical operations. Although this works satisfactorily in numerous cases, there remain nevertheless valuable stereoselective synthetic methods the description of which requires taking into account a more complex stereochemical knowledge. In order to solve this problem, we propose a general computer-oriented representation of the stereochemical aspect of transforms.

computer-aided organic synthesis design / stereochemistry / representation of synthetic methods / transform

Résumé — Représentation informatique des méthodes de synthèse stéréosélectives. Un système d'aide à la conception de plans de synthèse ne peut proposer des solutions pertinentes que s'il utilise une représentation informatique adéquate des méthodes de synthèse. Pour générer les précurseurs à partir d'une cible donnée, un tel système utilise les représentations rétrosynthétiques des méthodes de synthèse que l'on appelle transformations. Bien que la stéréochimie soit un facteur contribuant fortement à la complexité des problèmes de synthèse, seulement peu d'efforts ont été faits jusqu'à présent pour représenter l'aspect stéréochimique des transformations. Les approches connues sont basées exclusivement sur une connaissance élémentaire du cours stéréochimique des réactions, chaque transformation étant décrite en termes de quelques opérations stéréochimiques de base. Si ces approches donnent des résultats satisfaisants dans de nombreux cas, il reste néanmoins d'importantes méthodes de synthèses stéréosélectives dont la description nécessite l'utilisation d'une connaissance stéréochimique plus complexe. Pour remédier à ce problème, nous proposons une représentation informatique générale de l'aspect stéréochimique des transformations.

conception de plans de synthèse assistée par ordinateur / stéréochimie / représentation des méthodes de synthèse / transformation

Introduction

The development of computer programs to assist chemists in the planning of organic syntheses has now been going on for about 30 years. Essentially, two categories of programs have emerged [1]: (i) those that operate in retrosynthetic [2] direction and are dedicated to help in the design of organic multistep syntheses; (ii) those that work in synthetic direction and aim at predicting the outcome of chemical reactions. Our work is dedicated to the development of RESYN [3] which is the retrosynthetic analysis module of a synthesis design system that will also provide reasoning in the synthetic direction.

The empirical knowledge base of RESYN essentially consists of a transform [4] library which contains retrosynthetic representations of relevant synthetic methods. A synthetic method is a concept corresponding to either a class of chemical reactions or a combination of such classes which allows to create a significant structure. Above all, we need to represent the following knowledge associated with the transforms: (i) the modification of the structure of a given target molecule in order to generate the structure of the synthetic precursor(s); (ii) the situations that are (un)favorable to the application of a given transform.

While predicting the outcome of chemical reactions requires to account for their mechanism, in the context of synthesis design it is the net structural

* Correspondence and reprints

change effected by synthetic methods that must be taken into account. A transform is an operator which retrosynthetically modifies the computer-oriented representation of a chemical structure according to this net structural change. It is composed of the following primitive operators: addition/removal of atoms, creation/destruction/modification of bonds or stereogenic units [5].

In this article we shall focus on the representation of the stereochemical aspect of transforms. Despite the importance of stereochemistry for organic synthesis design, only few publications on this subject have come out. Ugi [6] and Wipke [7] have published their fundamental work already more than twenty years ago. Both methods can be used, in principle, for the representation of reactions in the synthetic as well as in the retrosynthetic direction. They exclusively take into account basic knowledge about the stereochemical course of reactions, for example, that an addition to a double bond proceeds in *syn* or *anti* fashion, that a substitution at an sp^3 center proceeds with inversion or retention of configuration, or that the course of a pericyclic reaction is suprafacial or antarafacial. Every reaction is described in terms of a few elementary stereochemical operations reflecting this basic knowledge. While many stereoselective organic reactions can be adequately described in these terms, there remain nevertheless numerous reactions the description of which requires taking into account a more complex stereochemical knowledge, for example, the stereocontrol exerted by the existence of a preferred transition state or by asymmetric induction.

There are two reasons why the known techniques do not always allow to adequately represent the net structural change effected by a given synthetic method. First, the exclusive use of basic stereochemical knowledge leaves room for ambiguities. For example, for a given regiochemistry there exist two possible *syn* additions to a double bond. Both Ugi's and Wipke's methods always consider all possible stereochemical outcomes of each elementary operation. Neither provides a means of control that would allow to resolve such stereochemical ambiguities when necessary. This prevents these approaches from representing the stereoselectivity of valuable synthetic methods. The second reason for the shortcomings of the known methods is that the elementary stereochemical operations provided by Ugi or Wipke do not cover all conceivable types of correlation of stereochemical features between a target and its precursor(s).

We propose a general computer-oriented representation of the stereochemical aspect of transforms which allows to deduce the correct stereogenic units in the precursor(s) from stereogenic units of the target molecule, whatever the stereochemical course of the reaction(s) which is (are) involved.

How to overcome the problems of the existing approaches

First of all, we must avoid the stereochemical ambiguities, that is, we must ensure that each application of a transform generates a unique precursor (or a unique set of precursors, if the target structure becomes fragmented).

What exactly does 'application of a transform to a target molecule' mean? The application of a transform is closely related to the transform selection process which in turn is related to the notion of 'structural objective'. Following a certain strategy, one can decide that some part of the target structure should be established or preserved. We define a structural objective as such an action (establish/preserve) on a given substructure. The goal of a synthetic method is to reach a structural objective which aims at establishing a substructure. For example, a synthetic method may allow to construct a certain type of polycyclic system, or to establish an assembly of several stereocenters in a certain relationship, or even create this assembly of stereocenters embedded in a polycyclic system. Although a structural objective globally affects a certain substructure, only some of the components of the substructure are actually modified. For example, the formation of a ring system requires creating only some of its bonds.

A synthetic method may allow one to attain several structural objectives concerning substructures of different synthetic complexity. Figure 1 shows some substructures that can be established using the aldol condensation, ranging from the basic substructure **3** over the cyclic substructure **4** to the *cis* fused bicyclic substructure **5**. Note that these substructures have a minimal substructure **1** in common which includes all the components modified by the synthetic method, and which we call the operand of the transform. A transform retrosynthetically specifies the corresponding structural changes with respect to the operand. For example, the transform corresponding to the aldol condensation is graphically represented by **1** ⇒ **2**.

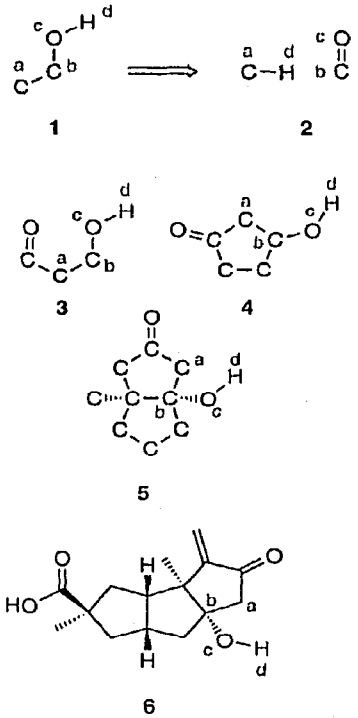

Fig 1

In order to modify a given target structure according to a certain transform, a synthesis design system must determine a valid one-to-one correspondence (bijection) between the atoms of the operand of the transform and those of a substructure of the target. This is accomplished as follows. Let our structural objective be to establish the *cis* fused bicyclic system (middle and right 5-membered rings) of target **6** [8]. The synthesis design system selects, via comparison of substructures, a transform capable of attaining this or an even more specific structural objective. In our example, because substructure **5** includes two *cis* fused 5-membered rings, and because target **6** includes substructure **5**, the transform corresponding to the aldol condensation will be selected. The one-to-one correspondence between the atoms of the target substructure and those of the operand are indicated by the lowercase letters in figure 1. Taking into account this mapping, the transform corresponding to the aldol condensation can be applied to target **6**, that is, the substructure of the target can be modified according to **1** ⇒ **2**.

Because each application of a transform to a target structure is to generate a unique precursor (or set of precursors), we admit only stereochemically unambiguous structural modifications in our transform descriptions. In contrast to this, the existing approaches describe transforms using inter alia ambiguous elementary stereochemical operations. This implies that one application of a transform may simultaneously generate several precursors, some of which may be stereochemically incorrect. As a consequence, the known methods fail to represent the stereoselectivity of many transforms.

If a transform has been selected to attain a structural objective for a given target structure, this means that the corresponding substructure of the target fulfills the necessary structural conditions for the transform to be applied. However, these conditions may not be sufficient. In general, for reasons related to chemical considerations or to the transform description we must additionally require that the structural context of the target substructure be not unfavorable for the transform application. This technique of specifying structural contexts will be illustrated below, where it is used to represent the *endo* selectivity of the Diels–Alder reaction.

Moreover, we introduce the concept of 'mapping of stereogenic units'. This means the creation of a stereogenic unit from another one by replacing the atoms of the former according to a given atom-to-atom mapping. In its generalized form, the mapping of stereogenic units allows one to create a stereogenic unit of any type (double bond, asymmetric carbon, etc) depending on another stereogenic unit. With this technique, it becomes possible to deduce stereogenic units of the precursor from those of the target, even if the correlation of these stereochemical features is too complex to be expressed by elementary stereochemical operations.

In the following we shall explain the techniques used in our approach with the help of examples illustrating the relevance and generality of these techniques. None of the corresponding transforms can be adequately represented using the existing methods.

Mapping of stereogenic units

Let us suppose that during the retrosynthetic analysis of a target molecule the modification of substructure **7** (fig 2) has been recognized as a strategic objective. We must therefore have at least one synthetic method at our disposal which allows to attain – perhaps indirectly – this objective. For example, the stereospecific hydroxylation of a double bond under the control of the configuration of a sulfoxide group (which is subsequently oxidized under the reaction conditions) is a known reaction [9]. Figure 2 shows that this reaction will produce **9** from educt **8**. One could therefore imagine to subsequently both hydrolyze the amide and take advantage of the reactivity of the sulfoxide in order to use this reaction for obtaining **7** from **10** via **7a**. (In the same manner, the stereospecific hydroxylation of **8′** leads to **9′**, and it should be possible to obtain stereoisomer **7′** from **10′**.) This leads to the definition of **7a** as a new objective.

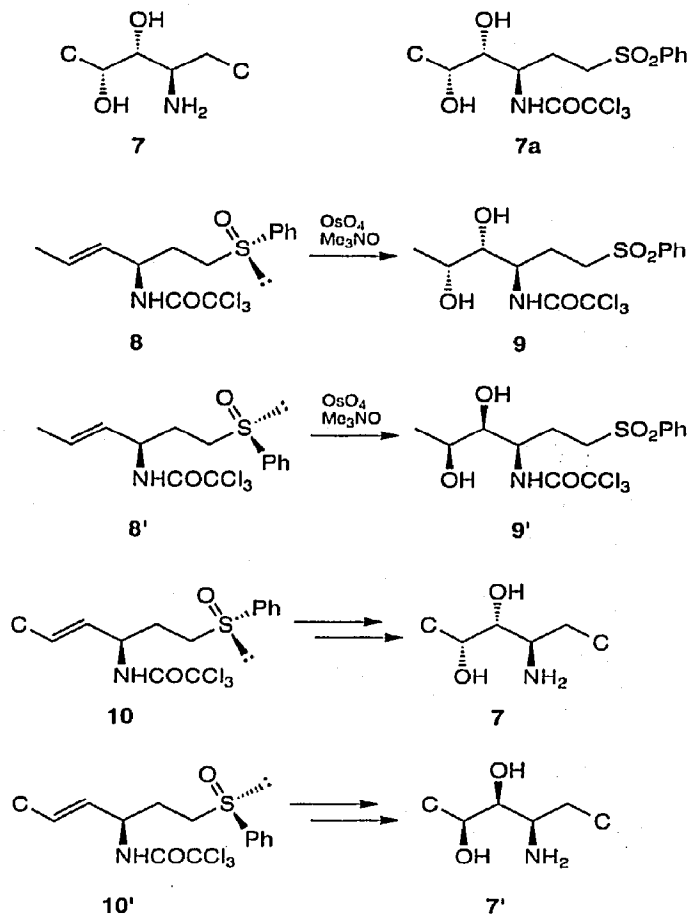

Fig 2

During the development of the retrosynthetic tree, our synthesis design system will therefore have to first select and then apply the transform corresponding to the stereospecific hydroxylation of olefins mentioned above. This transform is formulated in figure 3. It is

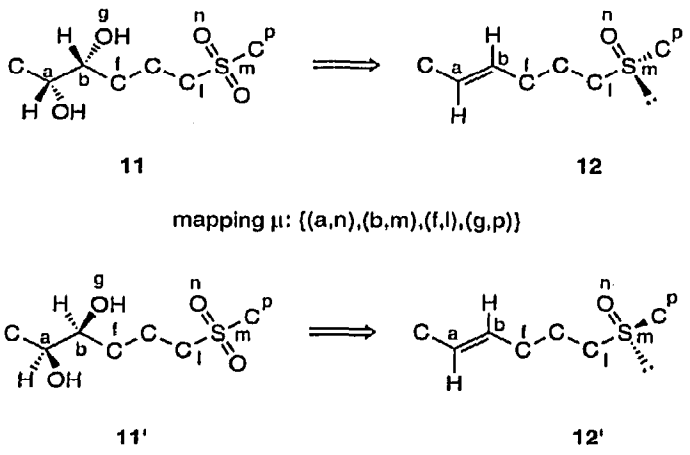

Fig 3

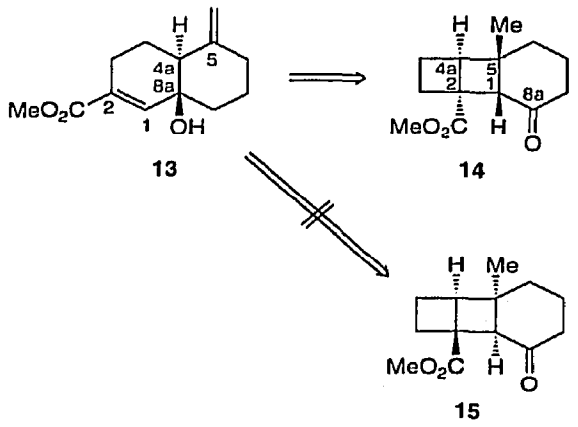

Fig 4

the operator which modifies either of the two enantiomorphic substructures **11** and **11′** in a given target structure in order to generate the structure of the corresponding precursor. Besides the constitutional modifications concerning the atoms and bonds, the stereochemical modifications that are to be made consist in removing the configuration of the two stereocenters a and b in **11** (**11′**), creating the conformation [10] of the double bond in **12** (**12′**), and in creating the configuration of the new stereocenter m in **12** (**12′**). The former two modifications can be achieved by an elementary stereochemical operation which is the *syn* elimination of two hydroxyl groups. In the latter case, however, the elementary stereochemical operations of the known methods do not allow to stereospecifically create the configuration of the new stereocenter. In fact, such an operation necessarily creates the two possible configurations of m, such that from a target molecule containing either **11** or **11′** the pair of precursors containing both **12** and **12′** is generated. In order to solve this problem, we need to describe the dependence that exists between the configurations of the stereocenters which are removed by the transform and the configuration of the stereocenter that is created. The notion of dependence is introduced by the mapping μ (fig 3) which expresses the correlation between the configuration of carbon b (more precisely, the subconfiguration of b which includes atoms a, b, f, and g) in **11** (**11′**) and that of sulfur m in **12** (**12′**). This allows to generate the configuration of m from a copy of this subconfiguration by replacing atom b by m, a by n, f by l, and g by p, according to mapping μ. (Note that we have made an arbitrary choice. Instead of the above subconfiguration of b and mapping μ, we could have chosen any pyramidal subconfiguration of either atom a or b, together with a suitable mapping.) Therefore, if we represent our transformation in this manner, its application generates exclusively the stereochemically correct precursor.

The following is another example of a transform where the stereogenic units in the precursor cannot be deduced from those in the target by means of elementary stereochemical operations. Target molecule **13** (fig 4) can be obtained from precursor **14** [13] by a tandem reaction in which the cleavage of the ring junctions 4a–2 and 5–1 is followed by an ene reaction.

If we would use Wipke's method [7] to describe the corresponding transform and apply it to **13**, we would wrongly obtain both **14** and **15** as precursors. Ugi's method [6] would additionally generate two further stereochemically incorrect precursors. Our corresponding transform uses the technique of mapping stereogenic units in order to deduce the configurations of atoms 1, 2, 4a, and 5 in the precursor from that of atom 4a (or 8a) in the target **13**, thus exclusively generating the correct precursor **14**.

Even in the case of reactions which follow a stereochemical course that is a priori simple, the known approaches may fail to adequately represent the corresponding transforms.

One example is the ring contraction of lactones via Ireland–Claisen rearrangement [14] (fig 5). The *Z* macrolactones **17** stereoselectively yield *cis* substituted rings **16**. (The case of the *E* macrolactones does not seem to have been studied extensively. In the known cases, a mixture of *cis* and *trans* products [14] or the *trans* product [15] has been obtained.)

According to either Ugi's or Wipke's method, application of the corresponding transform to target **16** would yield not only the correct precursor **17** but wrongly also **18**. (Both methods do work, however, fo

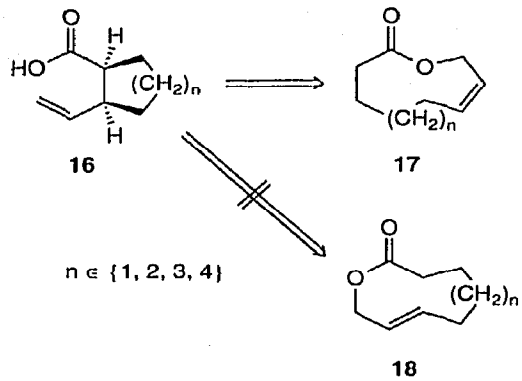

Fig 5

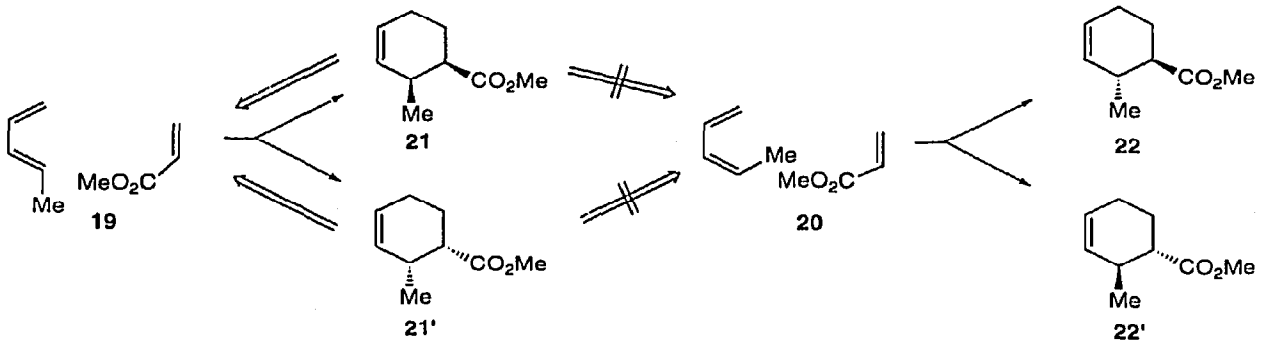

Fig 6

Claisen or Ireland–Claisen rearrangements of acyclic molecules. In this case, a given target may actually be obtained from two stereochemically different precursors.) In contrast to this, we deduce the conformation [10] of the double bond in the precursor from the relative stereochemistry of the substituted target ring, thus generating exclusively the correct precursor **17** from target **16**.

Use of structural contexts

The Diels–Alder reaction is another example of an important synthetic method which follows an a priori simple stereochemical course but which cannot be adequately represented by the existing approaches. This reaction proceeds suprafacially with respect to both diene and dienophile. Moreover, under kinetic control the intermolecular Diels–Alder reaction stereoselectively follows an *endo*-addition pathway, as a rule. Figure 6 shows a simple example [16]: the set of molecules **19** yields the enantiomeric cyclohexenes **21** and **21′**.

The methods of Ugi and Wipke merely allow to take into account the suprafacial course of the Diels–Alder reaction, but not the *endo* selectivity. Consequently, if we would apply the corresponding transforms to target molecule **21** (or **21′**), we would not only obtain the correct set of precursor molecules **19** but also **20**. However, the latter set of molecules would probably only show poor reactivity [17] and produce the cyclohexenes **22** and **22′** instead of **21** and **21′** [18].

We introduce the technique of specifying a structural context in order to represent the *endo* selectivity of the Diels–Alder transform, thereby avoiding these problems [19]. Figure 7 shows a graphical representation (**23** ⇒ **24**) of this transform. The necessary structural context for the transform application is given by the following requirement, concerning the one-to-one correspondence between the atoms of the cyclohexene substructure of a given target Diels–Alder adduct (including the atoms directly attached to the ring) and those of the operand **23**: the transform **23** ⇒ **24** can only be applied if the target substituents stemming from the dienophile are attached via atoms corresponding to atoms m and/or o of **23**. Figure 7 shows that exactly one such atom correspondence exists for each of the target molecules **21** and **21′**. Therefore, in either case the transform **23** ⇒ **24** can be applied to generate exclusively the correct set of precursor molecules **19**, that is, it takes into account both suprafacial *endo*-addition pathways.

Together, the transform **23** ⇒ **24** and the required structural context describe the stereochemical relationship between the stereogenic units in the target and the set of precursor molecules. The structural context required above adequately accounts for the *endo* selectivity in our simple example, where one of the ring atoms stemming from the dienophile is unsubstituted while the other one is monosubstituted. It is also valid when both of these atoms are monosubstituted with the substituents in *cis* relation. Should the two substituents be in *trans* relation, or should at least one of these atoms be disubstituted, priority rules are needed to determine the positioning of the substituents on positions m, n, o, and p. For example, we could preferably put onto positions m and/or o of **23** such substituents that are capable of secondary orbital interactions in the transition state of the Diels–Alder reaction [26, 27].

A similar problem arises in the case of the intramolecular Diels–Alder reaction. This synthetic method allows to obtain, for example, target molecule **25** (fig 8) from precursor **26** [28].

If the corresponding transform was represented using either Ugi's or Wipke's method, its application to the target **25** would not only generate the correct precursor **26**, but wrongly also **27**. However, *E* dienes react to give *trans* ring junctions [29]. In contrast to this, by specifying a structural context necessary for the transform application we can ensure that our intramolecular Diels–Alder transform exclusively generates the correct precursor. Using the same transform **23** ⇒ **24** as above (fig 7), we require for the one-to-one correspondence between the atoms of the cyclohexene substructure of the target (including the atoms directly attached to the ring) and those of **23**, that the atom belonging to the 5-membered ring and being attached in β position with respect to the cyclohexene double bond must correspond to either atom n or atom p. In the case of target **25** this implies a correspondence between atom β of **25** (fig 8) and atom p of **23** as well as between atom α of **25** and atom h of **23** (and atom h of **24** after application of the transform).

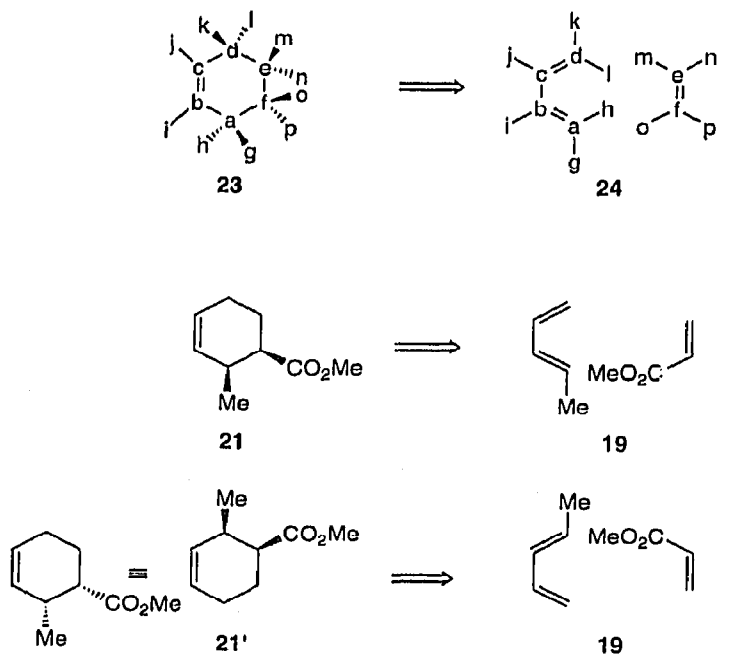

Fig 7

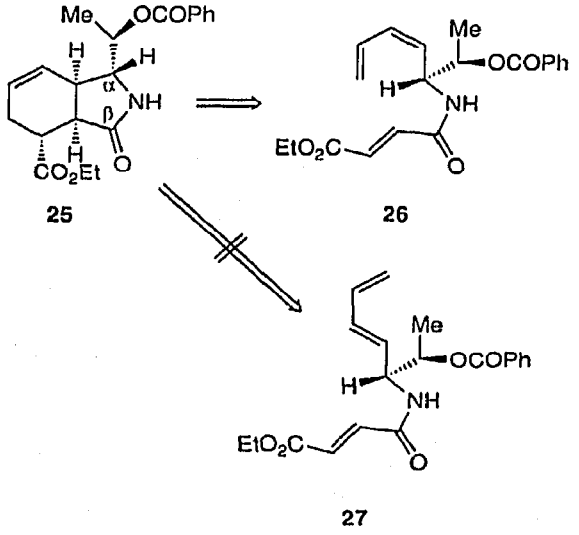

Fig 8

Formal transform description

In the previous sections, we have illustrated the essential ideas of our transform representation using the stereospecific hydroxylation of olefins and the Diels–Alder cycloaddition as examples. Here we give a comprehensive and formal description of these two transforms.

We have recently proposed a representation of molecular structure in which constitution, configuration and conformation of a chemical structure are represented independently from each other [11]. The configuration of an atom has been formally defined as the relative spatial arrangement of this atom and its neighbor atoms. Similarly, we have defined the conformation [10] of a bond between two atoms in terms of the relative spatial arrangement of these atoms and their neighbor atoms. The configuration and the conformation of a chemical structure have been defined as the set of atomic configurations and the set of bond conformations. Accordingly, the description of a transform consists of three mutually independent parts, and primitive set operations are used to describe constitutional as well as configurational and conformational changes.

Consider first the stereospecific hydroxylation transform (fig 3) which changes either structure **11** into structure **12** or structure **11′** into structure **12′**, depending on the target molecule. For the sake of simplicity, we shall develop our stereospecific transform considering the case **11** ⇒ **12**.

If AtCfg(*at*,*cs*) is the configuration of atom *at* in chemical structure *cs*, then Cfg(**11**) = {AtCfg(a,**11**), AtCfg(b,**11**)} is the configuration of structure **11** (fig 3) and Cfg(**12**) = {AtCfg(m,**12**)} that of structure **12**. We obviously obtain Cfg(**12**) from Cfg(**11**) through removal of AtCfg(a,**11**) and AtCfg(b,**11**) and addition of AtCfg(m,**12**).

Note that the stereospecificity of the transform requires us to deduce AtCfg(m,**12**) from either AtCfg(a,**11**) or AtCfg(b,**11**) via a mapping of stereogenic units. We arbitrarily choose AtCfg(b,**11**), or more precisely, $\mathrm{AtCfg(b,11)}|_{\{a,b,f,g\}}$, the subconfiguration of AtCfg(b,**11**) which takes into account only the relative spatial arrangement of the atoms in {a,b,f,g}. We generate AtCfg(m,**12**) from $\mathrm{AtCfg(b,11)}|_{\{a,b,f,g\}}$ by applying the mapping μ = {(a,n),(b,m),(f,l),(g,p)}, that is, by replacing b by m, a by n, f by l, and g by p (fig 3).

Let add(e,s) and remove(e,s) denote the sets which are obtained by addition and removal of element e to/from set s. Let furthermore map(e,m) denote the image of element e under mapping m. Then the configurational part of our transform case **11** ⇒ **12** is:

add(map(AtCfg(b,**11**)|$_{\{a,b,f,g\}}$,μ),
  remove(AtCfg(a,**11**),
    remove(AtCfg(b,**11**),Cfg(**11**)))).

In the preceding discussion, **11** and **12** may be consistently replaced by **11′** and **12′**. If we allow **X** to take either the value **11** or **11′**, the configurational part of our stereospecific transform is therefore:

add(map(AtCfg(b,**X**)|$_{\{a,b,f,g\}}$,μ),
  remove(AtCfg(a,**X**),
    remove(AtCfg(b,**X**),Cfg(**X**)))).

What about the changes in conformation [10]? In both **12** and **12′** rotation about the double bond between atoms a and b is hindered, whereas neither **11** nor **11′** has a double bond (fig 3). Therefore, the conformations of **11** and **11′** are Cfm(**11**) = Cfm(**11'**) = { }. Moreover, the conformation of the double bond in **12** is the same as in **12′**: if we denote this conformation by db and if BdCfm((*at1*,*at2*),*cs*) denotes the conformation of the bond between atoms *at1* and *at2* in chemical structure *cs*, we have db = BdCfm((a,b),**12**) = BdCfm((a,b),**12′**). Consequently, the conformations of **12** and **12′** are Cfm(**12**) = Cfm(**12′**) = {db}. If we allow **X** to take either the value **11** or **11′**, the conformational part of our stereospecific transform is therefore add(db,Cfm(**X**)).

Note that the application of this transform to target **9** (fig 2) generates precursor **8** whereas its application to target **9′** generates precursor **8′**, that is, the transform adequately represents the stereospecificity of the corresponding hydroxylation reaction.

Let us now take a look at the Diels–Alder transform (fig 7): the configuration of operand **23** includes four potentially asymmetric atoms, that is, Cfg(**23**) = {AtCfg(a,**23**), AtCfg(d,**23**), AtCfg(e,**23**), AtCfg(f,**23**)}, whereas **24** includes none, that is, Cfg(**24**) = { }. Thus, the following represents the configurational part of the Diels–Alder transform **23** ⇒ **24** (fig 7):

remove(AtCfg(f,**23**),
  remove(AtCfg(e,**23**),
    remove(AtCfg(d,**23**),
      remove(AtCfg(a,**23**),Cfg(**23**))))).

The conformations of **23** and **24** are Cfm(**23**) = {BdCfm((b,c),**23**)} and Cfm(**24**) = {BdCfm((a,b),**24**), BdCfm((c,d),**24**), BdCfm((e,f),**24**)}. Therefore, we represent the conformational part of the Diels–Alder transform **23** ⇒ **24** as follows:

add(BdCfm((e,f),**24**),
  add(BdCfm((c,d),**24**),
    add(BdCfm((a,b),**24**),
      remove(BdCfm((b,c),**23**),Cfm(**23**))))).

Note that these structure manipulation instructions represent both the intermolecular and intramolecular Diels–Alder transform. The *endo* selectivity of the intermolecular Diels–Alder reaction, for example, is not expressed within the transform itself. It is rather represented by requiring a certain structural context in the target structure for the transform to be applicable (see the discussion of the Diels–Alder transform in the preceding section) [30].

Conclusion

We have proposed a general computer-oriented representation of the stereochemical aspect of transforms. This work has been necessary because there exist valuable synthetic methods the stereoselectivity of which cannot be adequately represented using the existing approaches.

Our method overcomes the deficiencies of the known approaches as follows. First, we have introduced the technique of mapping of stereogenic units which enables us to create stereogenic units of the precursor from those of the target, even if the correlation of these stereochemical features is too complex to be expressed by elementary stereochemical operations. Second, each application of a transform to a target structure generates a unique precursor (or a unique set of precursors, if the target structure becomes fragmented). Third, the technique of specifying necessary stereochemical structural contexts is a means of controlling how a transform can be applied to a target, that is, it allows to preclude certain one-to-one correspondences between a substructure of the target and the operand of the transform. From this results a controlled one-by-one generation of precursors, which avoids stereochemically wrong precursors without missing correct solutions.

Although this article essentially presents the use of our method for the retrosynthetic representation of synthetic methods, it can also be used to represent chemical reactions in the synthetic direction. Currently, we are implementing and testing this stereochemical transform representation within the framework of the RESYN program for retrosynthetic analysis.

References and notes

1 Ott MA, Noordik JH, *Recl Trav Chim Pays-Bas* (1992) 111, 239
2 Corey EJ, *Pure Appl Chem* (1967) 14, 19
3 Vismara P, Régin JC, Quinqueton J, Py M, Laurenço C, Lapied L, in: *Actes des 12èmes Journées Internationales sur les systèmes experts et leurs applications*, EC2, Avignon, 1992, Vol 1, pp 305–318
4 Corey EJ, *Quart Rev Chem Soc* (1971) 25, 455
5 In contrast to this, the computer-oriented reaction representation for chemical reaction storage and retrieval systems requires a different approach. Such databases contain individual chemical reactions of which both educt and product sides are known. Database queries are (sub)structure-oriented, and therefore educt/product pairs adequately represent chemical reactions for this type of application.
6 Blair J, Gasteiger J, Gillespie C, Gillespie PD, Ugi I, *Tetrahedron* (1974) 30, 1845
7 Wipke WT, Dyott TM, *J Am Chem Soc* (1974) 96, 4825
8 Lansbury PT, Wang NY, Rhodes JE, *Tetrahedron Lett* (1972) 20, 2053
9 Hauser FM, Ellenberger SR, Clardy JC, Bass LS, *J Am Chem Soc* (1984) 106, 2458

10 We need definitions of configuration and conformation that are solely based on the relative spatial arrangement of the atoms in a chemical structure. Informally, we define an 'atomic configuration' as the relative spatial arrangement of an atom and its neighbor atoms, whereas the 'conformation of a bond' between two atoms is defined as the relative spatial arrangement of these atoms and their neighbor atoms. (Formal definitions of these concepts are given in [11].) Accordingly, we consider a double bond to be an element of molecular conformation. The configuration and the conformation of a chemical structure then are the set of atomic configurations and the set of bond conformations. These definitions correspond to IUPAC definitions E-1.4 (c) and E-1.7 (c) of configuration and conformation [12]. Unfortunately, adopting these definitions rather than IUPAC definitions E-1.4 (a) and E-1.7 (a) always leads to some confusion among chemists. However, the latter definitions are not viable for our purpose, because they are also based on energetical considerations. For a discussion of the related problems, see Appendix 1 on p 486 in [12].

11 Dietz A, *J Chem Inf Comput Sci* (1995) 35, 787

12 *IUPAC Nomenclature of Organic Chemistry, Sections A, B, C, D, E, F and H*, Pergamon Press, Oxford, 1979

13 Wender PA, Hubbs JC, *J Org Chem* (1980) 45, 365

14 Funk RL, Abelman MM, Munger JD, *Tetrahedron* (1986) 42, 2831

15 Enders D, Knopp M, Schiffers R, *Tetrahedron: Asymmetry* (1996) 7, 1847

16 Inukai T, Kojima T, *J Org Chem* (1967) 32, 869

17 O'Shea KE, Foote CS, *Tetrahedron Lett* (1990) 31, 841

18 Conversely, if we would apply the corresponding representations of the Diels–Alder reaction (in the synthetic direction) to the set of educt molecules **19** (fig 6), we would not only obtain the actual reaction products **21** and **21′** but also **22** and **22′**.

19 A similar method that has been introduced by one of us [20–23] uses predicates, so-called 'filters', in order to represent the stereochemical aspect of reactions within the framework of permutational isomerism [24, 25]. However, this representation is educt/product pair-oriented and does not describe the necessary structural modifications in order to generate the (set of) precursor(s) from a given target molecule.

20 Dietz A, *Ein Relationenkonzept zur Modellierung molekularer Systeme*, PhD Thesis, Technische Universität München, München, 1993

21 Ugi I, Bauer J, Bley K, Dengler A, Dietz A, Fontain E, Gruber B, Herges R, Knauer M, Reitsam K, Stein N, *Angew Chem Int Ed Engl* (1993) 32, 201

22 Ugi I, Bauer J, Blomberger C, Brandt J, Dietz A, Fontain E, Gruber B, von Scholley-Pfab A, Senff A, Stein N, *J Chem Inf Comput Sci* (1994) 34, 3

23 Dietz A, Gruber B, Ugi I, *MATCH* (1994) 31, 37

24 Ugi I, Marquarding D, Klusacek H, Gokel G, Gillespie P, *Angew Chem Int Ed Engl* (1970) 9, 703

25 Ugi I, Dugundji J, Kopp R, Marquarding D, *Perspectives in Theoretical Stereochemistry*, Springer, Berlin, 1984

26 Fleming I, *Frontier Orbitals and Organic Chemical Reactions*, John Wiley & Sons, London, 1976, pp 106–109

27 The synthetically directed stereochemical representation of the Diels–Alder reaction would be **24** ⇒ **23** (fig 7) with the following requirement to account for the *endo* selectivity: the atoms via which substituents are attached to the dienophile component must correspond to atom m and/or o of **24**. Application of this representation of the Diels–Alder reaction to the set of educt molecules **19** (fig 6) would exclusively generate the actual reaction products **21** and **21′**, in contrast to the existing methods.

28 Yoshioka M, Nakai H, Ohno M, *J Am Chem Soc* (1984) 106, 1133

29 Carruthers W, *Cycloaddition Reactions in Organic Synthesis*, Pergamon Press, Oxford, 1990, p 144 ff

30 Concerning the synthetically directed stereochemical representation of the Diels–Alder reaction (see also [27]) we can obtain it from the configurational and conformational parts of the transform if we mutually replace the functions *add* and *remove* and if we replace Cfg(**23**) by Cfg(**24**) as well as Cfm(**23**) by Cfm(**24**). In a manner similar as for the transform, the *endo* selectivity of the intermolecular Diels–Alder reaction is represented by requiring a certain structural context in the educt structures for the reaction representation to be applicable.