SYNSUP, Synthetic Route Design System

Sumitomo Chemical Co., Ltd.
Organic Synthesis Research Laboratory
Tetsuhiko TAKABATAKE

We have developed SYNSUP and the supporting systems for users to access SYNSUP through both Intranets and the Internet. We herein summarize the history and the current status of the SYNSUP service provided to chemists in our company and also some group companies.

This paper is translated from R&D Report, "SUMITOMO KAGAKU", vol. 2009-Ⅱ.

Introduction

Chemical and pharmaceutical companies carry out research into the development of new chemical compounds on a daily basis. If candidate compounds that have the targeted physical or biological properties are determined through molecular design and screening, they are manufactured and brought to market through investigations into methods for synthesis, evaluation of chemical, biological and physical properties, applications for patents, safety testing, industrialization, etc. Among this series of studies, synthetic routes have been investigated in the screening of compounds, in the preparation of evaluation samples, and in the industrialization of the processes. Usually, researchers first assume both or either the key reaction or the source material to be used based on their own knowledge, experience and intuition, and conceive of some candidate routes while referring to reaction examples obtained from searches of reaction databases and the literature, and carry out experiments. When reactions do not go as presumed or when associated problems arise, some other routes are conceived and experiments done. To obtain the best route in a short period, it is desirable to examine a synthesis map that covers as many possibilities as possible at the route planning stage. However, under time pressure, pre-laboratory studies are limited. In addition, it is virtually impossible even for excellent synthetic chemists to memorize vast knowledge such as named reactions and the reaction mechanisms and to keep up with reactions using new reagents or catalysts that are reported monthly in chemical journals.

Shouldn't it be possible to let computers arrange the large volume of reaction data, combine it systematically and create a synthetic route map? Having this aspiration, we have progressed in the development of the SYNSUP synthetic route design system. After looking back on the research history, we will give an overview of the current SYNSUP system and the user operating environment. This report was presented on behalf of people who have been involved in the project in the past both inside and outside Sumitomo Chemical.

Synthetic Route Design System History [1)–3)]

Researches on synthetic route design using computers began at universities in Europe and the United States in the 1960s. OCSS,[4)] which was announced by E. J. Corey and Todd Wipke of Harvard University, was the very first instance. LHASA[5)] was developed by them later. These were information oriented systems that carried out retrosynthesis (reverse synthesis) using transforms (reverse reactions) that are compiled based on analysis of known reactions. Malcolm Bersohn of the University of Toronto started to build a program around the same time and developed the first automatic synthetic design system with Ashmeed Esack.[6), 7)]

In the 1970s, Ivar Ugi and Johann Gasteiger reported on CICLOPS, which handled reactions by the recombination of bonds and lone pairs.[8)] After that, Gasteiger

developed EROS.[9] These were the pioneers of logic oriented systems. Chemical corporations in Europe focused on Wipke's system, brought together a consortium of corporations and carried out joint development of CASP.[10]

Once we entered the 1980s, the REACCS reaction database search system[11] was developed using technologies derived from information oriented synthetic design systems. ISIS, the successor system to it, is popular today.

With the coming of the 1990s, corporations in Europe and the United States moved their interests from synthetic route design to molecular design systems. On the other hand, in Japan Kimito Funatsu began development of AIPHOS[12].

Entering the 2000s, companies that advocate the utility of synthetic route design systems as a part of the workflow in pharmaceutical development appeared (ChemSpire[13] by Row 2 Technologies, Inc and ARChem Route Designer[14] by Simulated Biomolecular Systems, Inc.). Refer to **Fig. 1** for the changes in prominent systems.

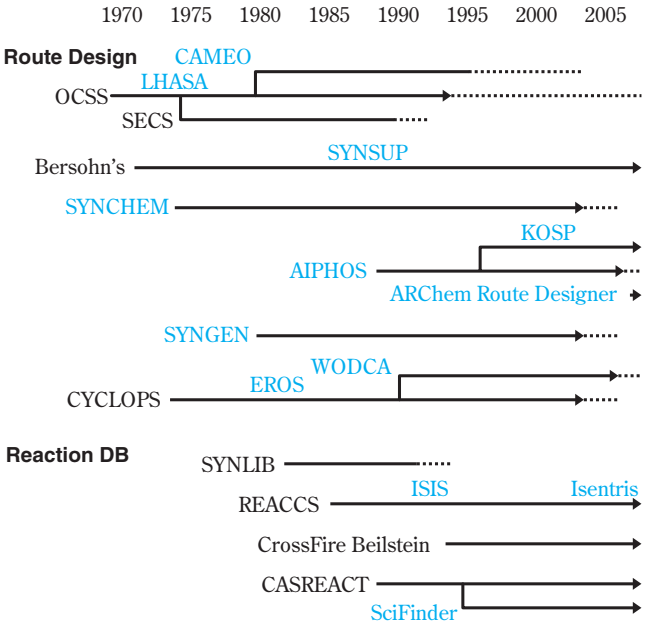


Fig. 1 Transition of major computer aides for organic synthesis

Work by Sumitomo Chemical [15), 16)]

After preliminary investigations into synthetic route design systems, Sumitomo Chemical started research in this field in early 1980s. Following researching literature, we had direct contact with researchers on some representative systems, and as a result, we introduced a logic oriented system (EROS) in 1983 and an information oriented system (Bersohn's program) in 1984. We found that the former was not suited to retrosynthesis of compounds with functional groups and subsequently introduced WODCA.[17] We started joint development with Bersohn for the SYNthesis SUmitomo Program (SYNSUP).[18] Besides them, we introduced a reaction prediction system (CAMEO)[19] and use it for pKa predictions. In addition, we evaluated CHIRON,[20] which proposes known chiral intermediates for complex asymmetric targets, and SYNGEN,[21] which carries out retrosynthesis using rules derived from reactions modeled based on reaction mechanisms. We introduced for synthetic chemists a reaction database REACCS that had later been replaced by ISIS.

For the purpose of the evaluation of SYNSUP, test runs were executed in 1990 for sample compounds jointly with a company that had introduced LHASA, and the results of the two programs were compared. In addition, we carried out a similar comparative examination with SYNCHEM[22] in 1998. In both cases, we confirmed that SYNSUP was by no means inferior to these typical information oriented programs. Therefore, we prepared a user-friendly operating environment and made it public to the entire company in 2000. After that, we have successively expanded into a user service for the companies of the Sumitomo Chemical Group. At present, we are running jobs around 700 times each year. The joint research with Bersohn ended at the end of 2007, and since then Bersohn and Sumitomo Chemical Co., Ltd. have both continued improving their program versions independently.

Overview of SYNSUP

1. Functional Concept

Fig. 2 shows the flow of processing in SYNSUP. Once a target compound is provided, the characteristics of the molecular structure are recognized, followed by acquisition of the reaction sites. For one reaction site taken up, the related reaction rule is tried to apply and a precursor (intermediate or reactant) is generated. At this time regioselectivity and chemoselectivity are checked. If there are no obstacles, a retroreaction is completed. If the precursor obtained is either that which is found in the available compound file or, that

which satisfies the constraints specified by the user, a single route is completed. If the precursor is not acceptable, the same procedure is applied to it by regarding it as a sub-target. If one route is completed, the system backtracks to the previous precursor, and the same operations are carried out for the next reaction site. In the end, the search stops when all of the reaction sites for the target compound and all of the intermediates have been tried. In the following, we will briefly describe the main components of the system.

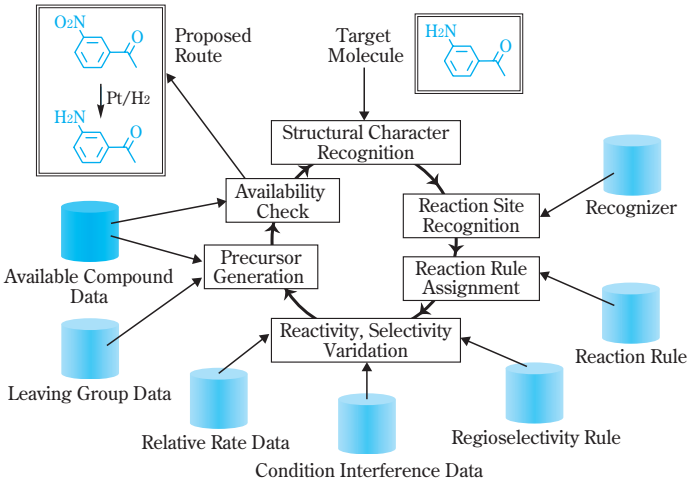
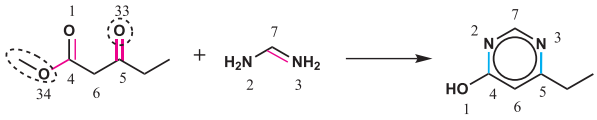


Fig. 2 Illustrative diagram of the search mechanism of SYNSUP

2. Reaction rules

The data required for carrying out a reverse reaction is called the reaction rule (see **Fig. 3**). The substructure composed by the atoms that are changed from before to after the reaction and the proximal atoms necessary for the reaction are called the reaction site, and what defines this is called the reaction site recognition code (recognizer). Recognizer numbers are used to refer to the reaction rules. The numbers of the substructures (these also being recognizers) that are changed in the precursors are listed for the purpose of checking possible competing reactions. Reaction condition (reagent) numbers that are used for checking chemoselectivity and a temperature range are specified. A typical yield is recorded from the literature. When the recognizer alone is insufficient to define the reaction site, separately prepared test numbers are specified. The former is a more basic rule, and the latter is used to distinguish between different reactions with similar transformations. When a leaving group is necessary, its information is specified. To generate

Reaction Scheme



Reaction Rule

```
REACTION NUMBER 5632 formamidine cyclization
SUBNUMBERS 2262[1] 136
NEXTSAME 0  NO PREVIOUS REACTION
CONDITIONS 20[2]     TEMP 25 25
YIELD 67
TESTS 169[3] 5 5080 169 6 5080
OPTIONS 1[4] 9 1 8 10[5] 2
MAKEBOND 2 4 3 5
BREAKBOND 5 33 5 33 4 34 1 4 3 7
REACTANTFUNGP 1 0 4 144 5 136 7 315
TITLE: A 4-hydroxypyrimidine from the condensation of a beta ketoester
and formamidine acetate with NaOMe in MeOH at room temp,
J. Heterocyclic Chem., 29,1369(1992),M. Butters,
Org. Lett., 7,3801(2005),J.C. Yoburn.
Two-step procedure: condensation of a beta ketoester and thioure,
followed by desulfurization by Ni/H2,
Org. Synth. Coll. Vol. iv 1963,638.
```

1) recognizer number
2) alkoxide
3) atom 5 must not be fused with aromatic ring
4) leaving group: 9 (unsaturated oxygen); 8 (methoxy)
5) aromatize the ring with atom 2

Recognizer

```
case 65: //phenolic OH, oxygen
  ...
  case 282: //N-C=N, central carbon
    makesb(mol,2262,// 4-hydroxypyrimidine
    1,i,1,       //sb[L][1] = i
    1,j,7,       //sb[L][7] = j
    33,1,7,4,    //sb[L][4] = NEAREST NBR OF ATOM 1 TO ATOM 7
    113,4,6,1,0,0,2,2,//ATOM 4:ARYL C, 2 HETERONBR
    3,4,7,2,     //sb[L][2] = COMMON NBR OF 4 AND 7
    113,2,7,1,0,0,0,0,//ATOM 2:UNSAT'D N, NO HETERONBRS
    57,2,2582,   //RINGINDEX MUST BE 2582
    13,4,1,2,6,  //sb[L][6] = A THIRD NBR OF 4 NOT 1 OR 2
    113,6,6,1,0,1,0,1,//ATOM 6:ARYL C,0-1 H, 0-1 HETERONBR
    7,6,4,5,     //sb[L][5] = OTHER RING NBR OF 6 BESIDES 4
    113,5,6,1,0,1,1,1,//ATOM 5:ARYL C,0-1 H, 1 HETERONBR
    3,5,7,3,     //sb[L][3] = COMMON NBR OF 5 AND 7
    113,3,7,1,0,0,0,0,//ATOM 3:UNSAT'D N, NO HETERONBRS
    0);          //SUBSTRUCTURE END
```

Fig. 3 An example of a reaction rule and the recognizer for it

reactant structures, the pairs of atoms involved in the required connections and disconnections are specified. When there are asymmetric centers or geometric isomerism that change or disappear from before to after the reaction, they are specified (these items are not found in the example in **Fig. 3**). The functional groups that change or disappear from before to after the reaction are specified. Finally, a brief description of the reaction and bibliographical information for the reference literature are noted.

The current library of reaction rules substantially covers typical reactions in Organic Synthesis, Collective Volumes I–X and other textbooks. Over the past 10 years, we have focused on heterocyclic synthesis

and added hundreds of useful rules. The total number of reaction rules has exceeded 5900.

3. Reaction site recognition

In case there exists no suitable recognizer when a reaction rule is to be registered, a new addition is made. The characteristics of atoms included in the reaction site are defined with the central atoms as the anchors, focusing on two structural characteristics at the reaction site. In the example of a recognizer shown in **Fig. 3**, a reaction site formed from a total of seven atoms of a pyrimidine ring and a hydroxy oxygen is defined with the oxygen of the aromatic hydroxy group and the carbon of the amidine as the anchors.

For the molecular structure given when SYNSUP is run, the functional groups, rings and stereochemical characteristics are recognized after canonicalization of the connection table for the molecule. Data on structural characteristics such as functional group numbers, ring numbers, asymmetric centers, geometrical isomerism and cis/trans relations of ring substituents is recorded in a structured data array. Allyl positions, benzyl positions, bridgehead positions, etc., that are related to reactivities are also recorded. Next, a search is made among the registered recognizers for applicable ones using the structural characteristic data in the molecule, and a list of substructures, which is indexed by the corresponding recognizer number, is created.

4. Reactivity and selectivity checks

Typical empirical rules are expressed as programming code. For example, when there are multiple reactive sites and they are not equivalent, byproducts are predicted, so the reaction is rejected. In rules for aromatic electrophilic substitution reactions, the reactivity of the reactant is estimated based on the electric and steric features of the substituents on the ring, and the applicability of the reaction rule is judged.

When it is thought that functional groups that exist outside of the reaction site currently being examined will react in the reaction condition checks, the reaction is rejected because byproducts will result. Particularly, checks are carried out in two stages. First, a check is done using reaction inhibition data and reaction competition data. The former is a combination of a reaction condition number and a recognizer number and the latter is a combination of a reaction condition number and two comparing recognizer numbers. Each of them is recorded along with the information on the literature on which it is based. When there exists no detailed data such as above mentioned, a check of chemical selectivity is carried out with reference to general-purpose reaction condition data. The interaction of each functional group and each reaction condition is arranged in a matrix of 150 structural characteristics × 180 reaction conditions.

5. Available compound file

Organic compounds of less than 72 atoms, excluding organic salts and polymers, which are handled by SYNSUP, are extracted from a database of chemicals catalog, i.e., a compound name, the supplier name and, if available, the price and the CAS NUMBER are recorded. Specifically, Chemicals Available for Purchase (CAP)[23] is used, and the data has been appropriately updated. Currently, 900,000 compounds are recorded in our available compound file, and of these, 30% have price data and 12% have CAS number.

The available compound file is consulted during the route search to determine the completion of a route. In addition, when a proposed route is output, it is used to provide chemicals data for available starting material and coreactants.

6. Search algorithm

If the reaction rules are applied unlimitedly to the reaction site list recognized in the target compound, it ends up with the generation of a huge number of routes. There are various mechanisms to prevent this. The reaction rules are classified into nine categories, and skeletal bond formation reactions are given higher priority (**Table 1**). This is because we will not get closer to completing the route if functional group interchange reactions are applied randomly. Degradation reactions, which are in the lowest position, are referenced only when the starting material is specified.

Pruning during the synthesis tree search is inevitable to cover as many desirable routes as possible. A complexity index of the target compound and all of the intermediates when generated is calculated for the number of functional groups, chiral centers, etc. When a route is completed, the complexity indices of the intermediates are thought of as a benchmark for the acceptable reactants at each step. In other words, when retroreaction is carried out and the complexity of one of the intermediates exceeds the minimum value so far in that depth of the tree, it is plausible to be an inefficient approach, so the search is broken off, and

Table 1 Classification of reactions for the reaction rule

priority	categories	descriptions about the reaction categories
1	specialr	enantioselective or makes more than 1 skeletal bond
2	hbuilder	diastereoselective or makes a skeletal C-Hetero or Hetero-Hetero bond
3	ccbuilder 1	makes a skeletal bond by losing at most 1 leaving group
4	ccbuilder 2	makes a skeletal bond by losing more than 1 leaving group
5	ufgi	makes a reactive group by functional group interchange (e.g., from allyl alcohol to allyl halide)
6	removeprotection	removes a protecting group
7	changfun	functional group interchange, usually trivial
8	removefun	removes a functional group by being replaced by H (e.g., deoxygenation of a ketone)
9	degrade	breaks off a carbon fragment (e.g., ozonolysis)

Reactions to introduce a protective group are classified as one of the categories with priority 1 to 7.

we backtrack and move to searching the next branch. The above is an overview of the exhaustive search algorithm.

For a complex target compound there are a large number of reaction sites and many synthesis steps are involved, which brings a severe combinatorial explosion. A selective search algorithm has been devised as a solution to this. After recognition of the structural characteristics of the target compound, strategically important bonds (key bonds) are identified based on the complexity centers in the molecule, and route search is done giving priority to reactions which generate a reaction site that includes one of the key bonds[24], the strategy of which is accordance with the concept of convergent synthesis. When no key bond formation reactions are approved due to problems in the reactivity or selectivity, removal of protection or functional group interchange reactions are hired to derive a precursor, to which a constructing reaction can be applied to cut one of the key bonds. It is noteworthy that the search is completed in a much shorter time than an exhaustive search.

7. Execution options

When the target compound is given and execution started, SYNSUP carries out synthetic route searches without the intervention of the user, so stopping rules for the search are necessary. Examples of execution options (search constraints) are given in **Table 2**. The most basic options are STEPLIMIT, which shows the upper limit for the number of steps, and CATALOG for the usage of the available compound file. Due to the default setting of STEPLIMIT 2 and CATALOG 2 (the starting material must be an available compound), execution is possible even if the user does not specify any options. When no route is proposed, STEPLIMIT is automatically incremented, and the program executed again. If the user specifies a STEPLIMIT that is unnecessarily large, it sets off redundant route searches with excessive execution time. In addition, desirable routes with shorter steps may not be proposed. The automatic setting of STEPLIMIT is extremely effective for preventing this sort of misuse.

Table 2 Typical Execution Options for SYNSUP

STEP LIMIT	maximum number of steps acceptable in a route
CATALOG	1 : cataloged compound is treated as available, 2 : only cataloged compound is acceptable as the starting material, 3 : the starting material and all coreactants in a route have to both be cataloged compound and satisfy other options specified
RING LIMIT	number of rings acceptable in the starting material
ATOM LIMIT	number of atoms acceptable in the starting material
FUNCTIONAL GROUP LIMIT	number of functional groups acceptable in the starting material
AROSUBST LIMIT	number of aromatic substituents acceptable in the starting material
NO_LG_VARIATIONS	route variations with only the difference of leaving groups to be rejected
STARTING MATERIAL	exact starting material to be used
ASYMMETRIC ONLY	only enantio- or diastereo-selective reactions to be applied
INDUSTRIAL	only industrially applicable reactions to be used
OMIT	omit the use of specified reactions

User Operating Environment

1. User interface

When it was first made public in 1996, the user operating environment for SYNSUP was a system where

each user using their PC logged in to a UNIX server and started the graphical user interface. However, access from locations outside of the research facility was not practical because the communication band was narrow. Therefore, a new Chemical Memory Bank edit (CMBedit) user interface program that can run on a PC was developed.

CMBedit is made up of a Molecule Editor that draws the structure of a target compound and a Synthesis Viewer that displays proposed routes. After using a template and pencil tool to draw the molecular structure (**Fig. 4**) in the former, the execution options such as STEPLIMIT and CATALOG are specified in a "standard options" tab (**Fig. 5**). **Fig. 6** shows the results when the program is executed with AROSUBST LIMIT (upper limit for the number of substituents on an aromatic ring) 1 specified. This is one of the four route display methods, the "Route map display." The target compound is drawn in the upper right, and the starting material is at the left edge of the reaction scheme. The blue boxes surrounding the compounds indicate that they are purchasable chemicals. If you double-click on the structural formula for one, the chemical's data is displayed in a pop-up window (**Fig. 7**). If you double-click on an arrow in a reaction equation, a brief description of the reaction and the bibliographical information from the reference literature are displayed in the same manner (**Fig. 8**). There is a function for hiding unwanted routes while browsing the routes and printing only the routes you want.

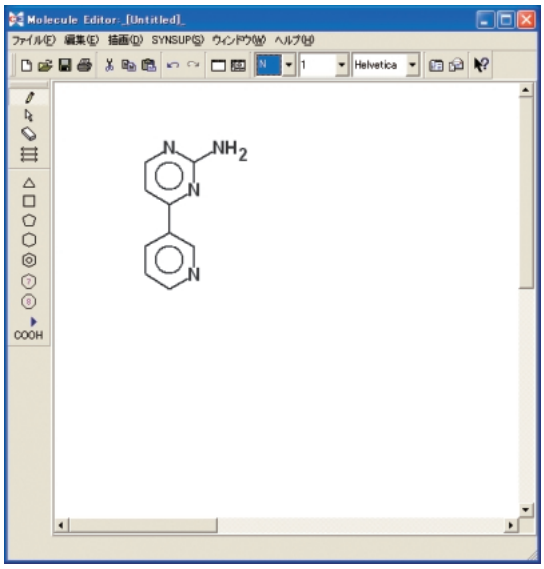


Fig. 4 CMBedit –Molecule editor

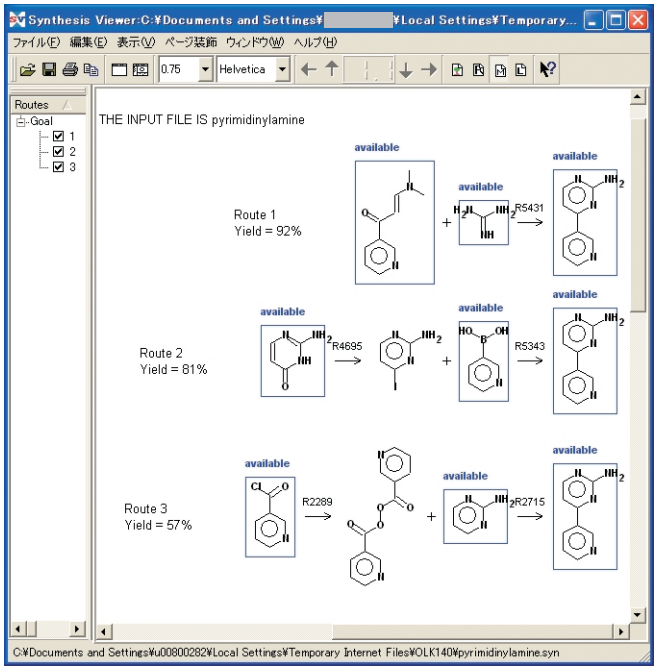


Fig. 6 CMBedit –Synthesis viewer–Map view

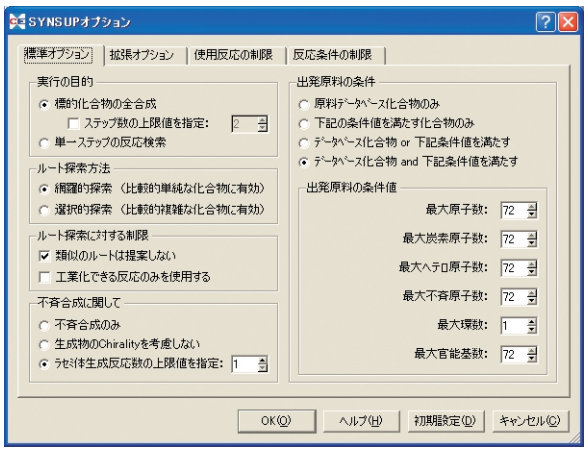


Fig. 5 CMBedit –Molecule editor–Option settings dialog box

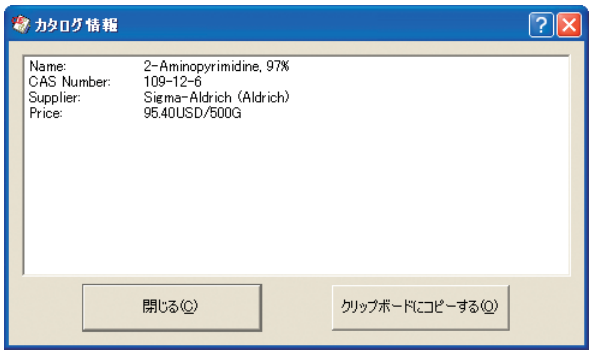


Fig. 7 CMBedit –Synthesis viewer–Catalog info

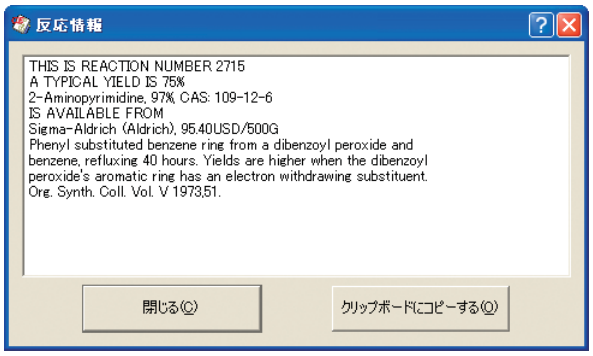


Fig. 8 CMBedit –Synthesis viewer–Reaction info

2. E-mail execution system

This is a system where a SYNSUP input file that a user has created on a PC is sent to the SYNSUP server by e-mail and the results of the execution of SYNSUP on the server returned to the user by e-mail (**Fig. 9**). We developed a set of programs that deals with the input data received as an email attachment from the user, that hands over it to the job execution system, and that returns an output file to the user by e-mail when the execution is completed, and we combined them with the TORQUE open-source batch job management system.[25)]

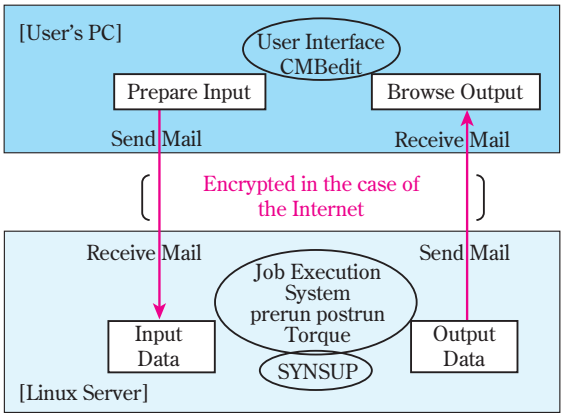


Fig. 9 SYNSUP Job execution system utilizing e-mail for input/output data transfer

3. Release to users

Having a practical user operating environment achieved with the combination of CMBedit and the e-mail execution system enabled us to extend the range of users in 2000. Starting with the Organic Synthesis Research Laboratory and the Agricultural Chemicals Research Laboratory, the program was released to other research laboratories and group companies capable of connecting to the Sumitomo Chemical LAN. Furthermore, we established a SYNSUP web site to provide the manual, FAQ and other reference materials as well as user registration and CMBedit installation services. In addition, we are operating an executive logging system for the purpose of user support and keeping the statistical record on the use of the program.

As for the use by group companies that cannot connect to the Sumitomo Chemical LAN, there was concern about data security for e-mail transmission over the Internet. By implementing an encryption function to both CMBedit and the e-mail execution system we have released the user operation of SYNSUP for group companies in 2008.

Example of SYNSUP Execution

Since SYNSUP is a system that depends on the reaction rules stored so far, it is unable to propose all feasible routes, but it can help the users select some optimal synthetic routes by supplementing what might be overlooked by them by indicating a number of synthetic variations. We will present some typical execution examples, though these are not ones that have actually contributed to research or industrialization.

1) Allethrolone[26)]

This is an intermediate for synthetic pyrethroids, which is one of Sumitomo Chemical's major products. As a result of execution with the constraints of STEPLIMIT 3 and CATALOG 2, 23 routes were proposed. **Fig. 10** shows a screen displaying only two representative routes that are publicly known. While most of them are natural-product-type synthethes such as the Route 4 where a ring is formed using the aldol reaction, Route 17 employs the furan-carbinol fragmentation reaction that is Sumitomo Chemical's own technology. We have obtained a possible rearrangement reaction, the reaction site of which was concealed at first glance. This is a great merit of using a computer system.

2) A new insulin-like growth factor 1 receptor inhibitor (amine part)

A test run was carried out for a compound recently reported in the literature. The final target compound is the one which is derived from the compound 1 in **Scheme 1**[27)] via amidation and deprotection. Taking

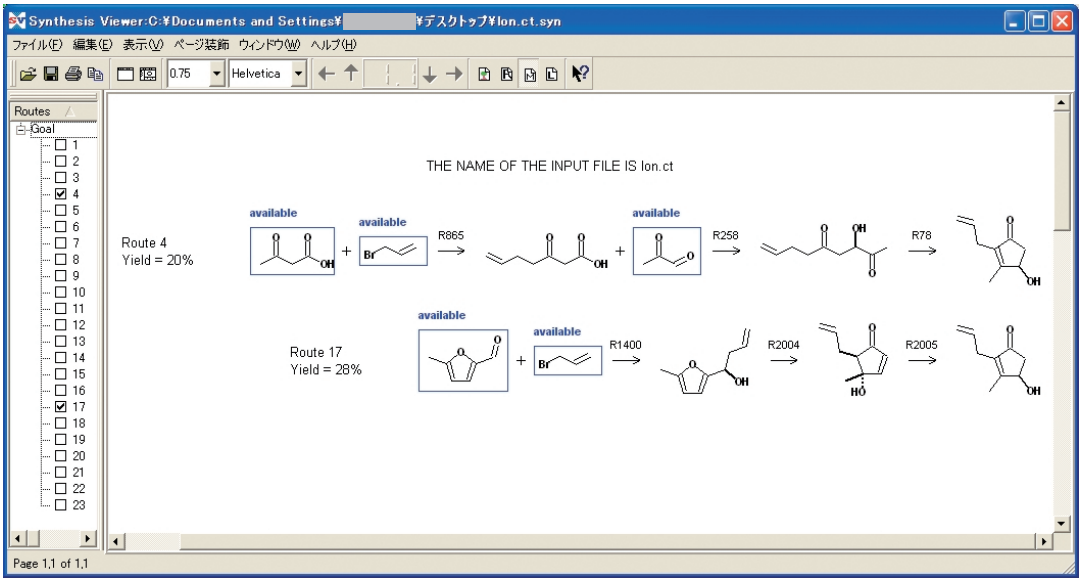
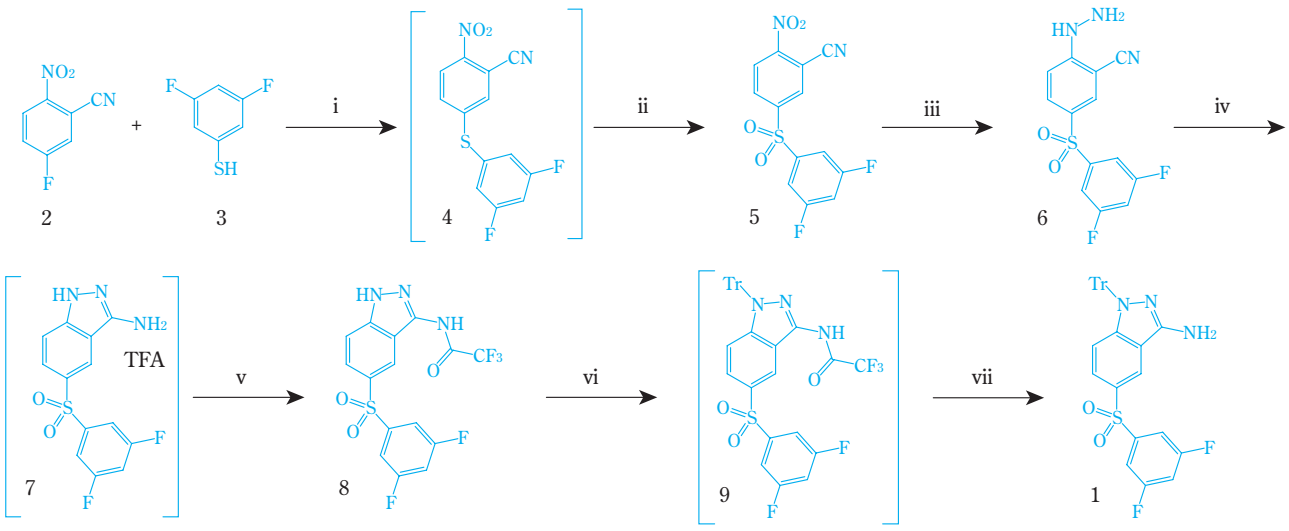


Fig. 10 Execution example 1: allethrolone

compound 7 which is 1 without the trityl group as the target, the execution was done with the constraints of STEPLIMIT 3 and CATALOG 2. There were more than 100 routes, in which fluorine substitution reaction of aromatic chlorides was used frequently. However we were able to narrow it down to 25 routes by adding an execution option of OMIT (do not use the specified reaction rules) with the number of fluorination reactions and resubmitting the job again. **Fig. 11** shows five routes where the rest were hidden due to their resemblance.

Routes 1, 8, 20, and 21 start with commercial chemicals having an aminoindazole skeleton. When screening is the purpose, these are probably routes to be considered. Though, all of the routes will require examination of the use of proper protecting groups. Route 3 is the same route as in the literature if we eliminate the absence or presence of protecting groups. When there is an undesirable reaction as in this example, we can narrow down the routes by re-executing the program using OMIT. In the nature of our tree pruning mechanism with some execution options added, it is possible to get different routes that were not seen in the first execution.



Reagents and conditions: i) EtOAc, DIPEA (1.05 equiv), 0 C to rt, 0.5 h; ii) MeCN, water, Oxone (2.5 equiv), 40 C to rt, 24; iii) THF, N_2H_4, 35%, rt; iv) THF, TFA, 5 C, 15 min; v) DCM, TFAA; vi) DCM, TrCl, TEA, rt, 4h; vii) MeOH, TEA, refluxed, 5h.

Scheme 1 Synthesis of a new insulin-like growth factor 1 receptor inhibitor (amine part) [27]

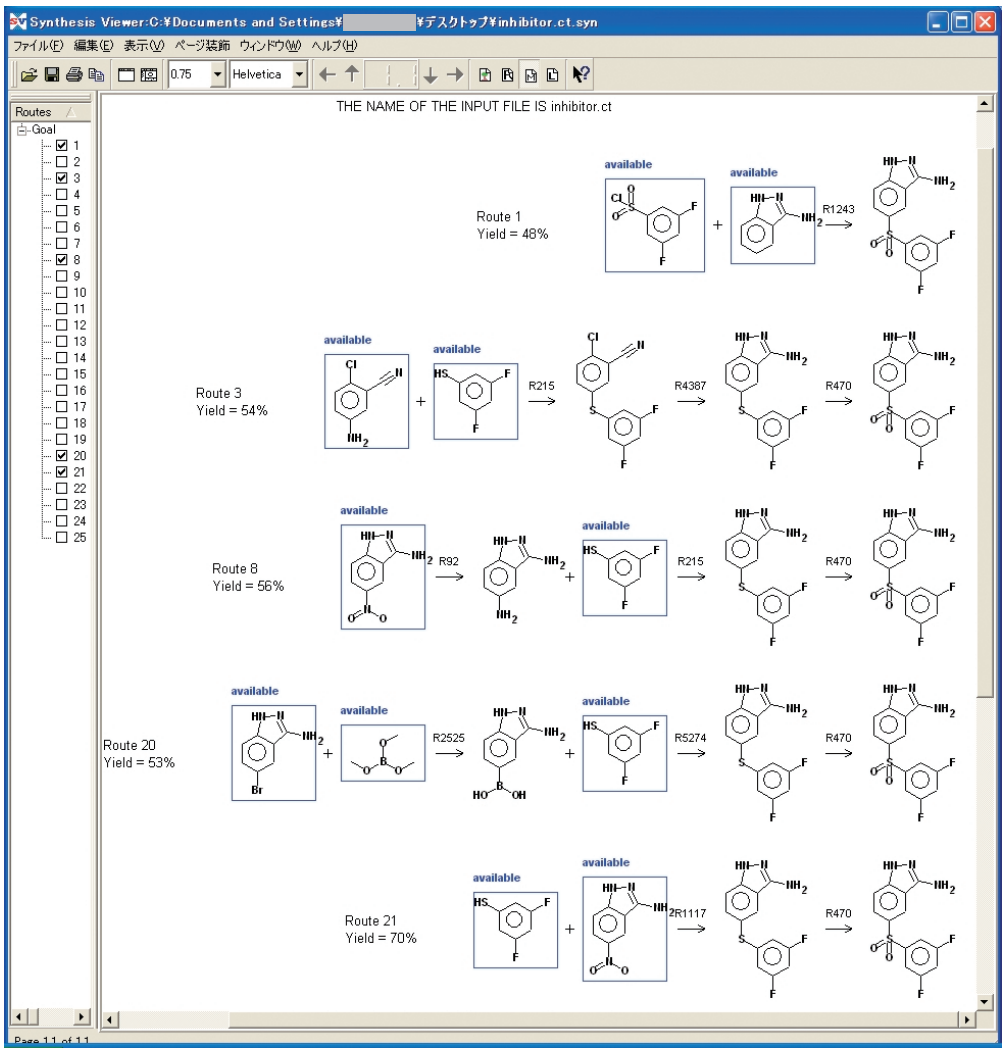


Fig. 11 Execution example 2: an inhibitor

A workstation with two X5460 Quad-Core Intel Xeon processors (3.16 GHz) was used for these runs. In the example in 2), the running time was one minute. With STEPLIMIT 3, a typical example completes within several minutes. However, the example in 1) required 1.5 hours. Compared with 2), it first appears to be a simple structure, but in fact there are more reaction sites and more applicable reaction routes than in 2).

Problems for the Future

We conducted a user survey in 2006. 30% of those responding selected that SYNSUP "is a useful information tool." When we include in the number of responses that it is "useful as a system supplementing other tools," almost every one evaluated it positively. The trend was for higher evaluations with a greater number of uses. On the other hand, the following were indicated as requests for improvements in SYNSUP.

- Often inappropriate reactions are used in the proposed routes.
- Many similar routes are output.
- Recent (particularly the 1990s and later) reactions are insufficient.

To improve the user satisfaction with SYNSUP, it is important to cover as many possible routes as possible and at the same time reduce unsuitable routes by as many as possible.

First of all, we must expand the library of reaction rules. At present, reaction rules are created in an interactive procedure using reaction rule creating tools after selecting a useful reaction from the literature. We are obtaining the cooperation of specialists outside of the company for selecting reactions. Apart from SYNSUP, there are a few systems that can create reaction rules automatically from commercial reaction databases (for example, ARChem Route Designer,[14] SYNCHEM,[28] and AIPHOS[29]). It is very attractive that a large

amount of reaction rules are generated without human assistance. However, we have to consider that there are many similar reaction examples in the reaction databases, and in addition, it is rare but there is some erroneous data included. The problem is how to collect only useful reactions when making reaction rules and/or select optimal reaction rules when executing the program. Furthermore, we must also consider the license cost for using commercial reaction databases. At present, we think that moving forward with having specialists select useful reactions from the literature and automating reaction rule creating tools is the most practical.

Second, it is important to improve the accuracy of side reaction checks. We must expand the reaction inhibition and competitive reaction data. However, even when we take up the reaction conditions that are known from the literature, the reaction rates often vary largely corresponding to the differences in the reactant structures, so it is not feasible to accurately determine the range of application at the time of creating a reaction rule. Therefore, every time any problems of selectivity appear in the proposed routes, the recognizers are reconsidered. Or, reaction inhibition and competitive reaction data have to be registered so that the range of applications is adjusted. It is desirable to have user comments to know problematic points in the routes proposed by SYNSUP. We hope to develop a system for obtaining feedback on unsuitable routes without burdening the users.

Conclusion

Synthetic route design is a challenge that essentially requires complex thoughts. For example, for structure identification of an unknown compound, there is just one solution, but in the case of synthetic routes, many solutions are possible. Unless suitable pruning is done on the synthesis tree, a large number of routes are output. On the other hand, to cover new or unconventional routes it is necessary to keep collecting knowledge on new findings about reactions that are updated daily in the chemical journals and patents. Furthermore, there are different evaluation criteria for the best route according to the purpose. That is, for screening compounds it is desirable to get routes with the fewest steps that combine reactions with a high level of reliability. On the other hand, routes for investigations into industrialization must be selected comprehensively from the standpoint of factors such as material costs, facilities costs and safety.

The history of synthetic route design systems exceeds 40 years, but we are still not at a stage where such commercial software has disseminated widely. For a synthetic route design system to become really popular on site for organic synthesis, it must be able to always propose routes that satisfy the researchers' expectation. This means, we must continue learning synthesis strategies from experts in synthesis and books to integrate the algorithms into the system along with collecting useful reactions from the literature.

At universities and corporate research laboratories today, the concern is moving from total synthesis of natural products to design of compounds that have new and useful functions and development of the applications. However, once the target compound has been determined, it is still necessary to design synthetic routes for establishing economical manufacturing processes. Along with the use of SciFinder, CrossFire Beilstein and ISIS, we expect that as the more SYNSUP evolves, the more synthetic chemists will take the advantage of the SYNSUP system for making synthetic route maps.

Acknowledgments

We would like to express our gratitude to Professor Emeritus Malcolm Bersohn (University of Toronto) who guided us for 24 years in joint development of the synthetic route design system on which this research is based.

References

1) W. L. Chen, *J. Chem. Inf. Model.*, **46**, 2230 (2006).
2) J. Gasteiger, Ed. "Handbook of Chemoinformatics", Wiley-VCH: Weinheim, Germany (2003).
3) T. Takabatake, H. Yamachika, "Yuki Gosei Kagaku", A. Oku, eds., Ouyou Kagaku Kouza 4, Asakura Shoten (1997), p. 209.
4) E. J. Corey and W. T. Wipke, *Science* **166**, 178 (1969).
5) E. J. Corey, W. T. Wipke, R. D. Cramer, III and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 421 (1972).
6) M. Bersohn, *Bull. Jpn. Chem. Soc.*, **45**, 1897 (1972).
7) M. Bersohn and A. Esack, *Computers & Chemistry*, **1** (2), 103 (1977).
8) J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie

and I. Ugi, "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann and E. Hyde, Eds., Wiley, New York (1974), p. 129.
9) J. Gasteiger and C. Jochum, *Topics Curr. Chem.*, **74**, 93 (1978).
10) P. Gund, E. J. J. Grabowski, D. R. Hoff, G. M. Smith, J. D. Andose, J. B. Rhodes and W. T. Wipke, *J. Chem. Inf. Comput. Sci.*, **20**, 88 (1980).
11) S. E. French, *CHEMTEC*, February, **1987**, 111.
12) K. Funatsu, S. Sasaki, "AIPHOS –Konpyuta niyoru Yuki Gosei Sekkei", Kyoritsu Shuppan (1994).
13) http://www.row2technologies.com/
14) J. Law, Z. Zsoldos, A. Simon, D. Reid, Y. Liu, S.Y. Khew, A.P. Johnson, S. Major, R. A. Wade and H. Y. Ando, *J. Chem. Inf. Model.*, **49**, 593 (2009).
15) M. Yoshida, T. Takabatake, M. Ishida, *SUMITOMO KAGAKU*, **1990-II**, 39 (1990).
16) T. Takabatake, T. Takemura, I. Dohgane, *SUMITOMO KAGAKU*, **1994-I**, 69 (1990).
17) W.-D. Ihlenfeldt, and J. Gasteiger, *Angew. Chem., Int. Ed. Engl.*, **34**, 2613 (1995).
18) M. Takahashi, I. Dogane, M. Yoshida, H. Yamachika, T. Takabatake, and M. Bersohn, *J. Chem. Inf. Comput. Sci.*, **30**, 436 (1990).
19) T. D. Salatin, and W. L. Jorgensen, *J. Org. Chem.*, **45**, 2043 (1980).
20) S. Hanessian, J. Franco, and B. Larouche, *Pure Appl. Chem.*, **62**, 1887 (1990).
21) J. B. Hendrickson, *Angew. Chem., Int. Ed. Engl.*, **29**, 1286 (1990).
22) H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, and J. E. Searleman, *Science*, **197**, 1041 (1977).
23) CAP: Chemicals Available for Purchase, Accelrys Software Inc.
24) A. Tanaka, T. Kawai, T. Takabatake, H. Okamoto, and M. Bersohn, *J. Comput. Aided Chem.*, **10**, 104 (2009).
25) TORQUE: http://www.clusterresources.com/products/torque-resource-manager.php
26) I. Dohgane, H. Yamachika, M. Minai, *J. Synth. Org. Chem Jpn.* **41**, 896 (1983).
27) I. Candiani, G. D'Arasmo, F. Heidempergher, and A. Tomasi, *Org. Process Res. Dev.*, **13**, 456 (2009).
28) H. Gelernter, J.R. Rose, and C. H. Chen, *J. Chem. Inf. Comput. Sci.*, **39**, 316(1990).
29) K. Satoh, and K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **39**, 316 (1999).

PROFILE

Tetsuhiko Takabatake

Sumitomo Chemical Co., Ltd.
Organic Synthesis Research Laboratory
Senior Research Associate