1. はじめに

化学は新しい物質を作ること、または物質の新しい作り方を研究することを中心とした学問ともいえる。したがって合成ルートに関するデータ、知識の体系化とその活用の重要性については改めて述べる必要はないであろう。このシリーズは計算機の利用を前提とした話であるが、一般的に化学の研究開発において必要なデータ、情報は必ずしも計算機処理には適していない。その理由はさまざまなものがあるが、物質自身とその変化に関する知識の不十分さがその主たるものであり、一方、情報処理の技術、理論がこのようなデータ、情報に十分に対応できる段階に達していないことにもよる。

合成は幸いなことに、数値情報に次いで計算機処理に適した対象であるから、ミニコンピュータの出現した時、すなわち 1960 年代の後半に合成経路の最適化の研究が始められた。最も有名なものはハーバード大学の E. J. Corey¹⁾ の LHASA (Logic and Heuristics Applied to Synthetic Analysis) であり、パチュリアルコールの 3 ないし 4 段の合成経路の解析を行ない、いくつかの興味あるルートが示された。このことは計算機が合成研究にとって、有効な道具になり得ること、合成に関するデータ、知識の入力、処理法の適切さを示したものであり、この分野に対する基盤となる画期的な一連の研究である。Corey の成果はその手法が以後踏襲され、計算機技術の進歩とともにその成果が多様な形で展開されたことは以下に述べることから伺えるはずである。

わが国においては化学研究の水準に比して、計算機

カット=ハーバード大学の Corey が開発した有名な合成ルート探索システム LHASA の構成図であり、基本的には現在でも通用するものである。

Retrieval of Chemical Reaction Information

合成ルートの検索

藤原 譲 Yuzuru Fujiwara, 筑波大学電子情報工學系

支援による分子設計は一部の分野を除いてきわめて普及度が低い。しかも欧米では薬品設計と並んで合成設計が実用化されているのに対し、反応設計は薬品設計に対しても出遅れているので、分子設計内部の分野の差も特異な現象を示している。ここでは合成ルートの最適設定のための手法の現状と問題点の概要を述べることにする。

2. 合成経路設計システムの現状

有機合成においては、官能基の反応を単位として、対象の化合物からの反応、またはそれを生成する前駆化合物を推定して、これを繰り返すことによって最終生成物、または出発物質とその間の経路を設定するのが有機化学者のとる方法であり、計算機上でも同様のことを行なわせることを試みるのであるが、システムによりその実現方法が異なるので代表的な場合について述べる。

2-1 LHASA

Corey は後に知識工学の主流となった。プロダクションルールすなわち IF (条件), THEN (処理 1), ELSE (処理 2) の形で素反応を記述する方式をとっている。合成経路、反応の逆をそれぞれ Antithesis, Transform と名づけて、目標化合物に対して出発物質とその経路を求めることを目的としたシステムである。一般的に反応は化合物間の変換に対応するので何

藤原 譲 筑波大学電子情報工學系 (305 茨城県新治郡桜村天王台 1-1-1) 教授・理博
昭和 32 年東大工学部応用物理学学科卒。32 年クラレ入社。51 年筑波大教授。32~42 年高分子 NMR, ESR の研究。42~50 年高分子統計、ラボオートメーションの研究。51~現在、情報学、データベース、知識工学の研究に従事。
<趣味>囲碁、旅行

らかの制約を付さなければ無限の可能性を順次掃引することになるため、探索の方向づけが必要でこれを戦略とよんでいる。反応段数のみでなく実際に即した戦略として3種のをシステムに組み入れている。第1は部分構造が反応規則を適用できる官能基であるかどうか調べ、適用不能のときは官能基の付加や交換、またはそれらの組み合わせを図ることである。次は目標化合物の単純化のために切断する結合の選択を戦略的に行なうもので、切断されたものになるべく小さくなることを基本として第1の戦略である官能基の修飾機能も組み合わせて用いられる。さらに包括戦略ともいべきものがあり、これは一連の多段反応の全体を包括する探索表を用意しておいて最大20段までの定型的逆合成をまとめて扱うことで探索を著しく単純化するのみならず、実際の反応手順、技術との良好な対応が得られる。

これらの戦略と個々の規則のみでは、段数の多い反応の場合、必ずしも意味のないものをも含めて処理することになるので LHASA では評価を点数の形で与え、実用性をチェックしている。

またカットにも示したように化合物や反応の情報はグラフィック表示や、入力を行なうことによって化学者にとって非常に便利なものとなり、当初より必須の機能として付けてある。

Corey の弟子でありその後 du Pont で分子設計システム TRIBBLE 実用化の中心になった Pensak²⁾ の合成ルート探索の例を図1に示す。反応は立体化学を含めてルール化されているのがわかる。

また Kaufman らの開発した PASCOP (Programme d'Aide à la Synthèse des Composés Organo Phosphorés) も LHASA の系統のシステムであるが名前の通り、含リン複素環化合物を対象とするシステ

図1 ヨウ素ラクトン化の変換系列

	1	9	10	2	11	12	3	4	5	6	7	8
	C	H	H	C	H	H	H	H	*	*	*	*
G ₁	1C	1	1	2								
	9H	1										
	10H	1										
G ₂	2C	2			1	1						
	11H				1							
	12H			1								
G ₃	3H						1					
G ₄	4H						1					
G ₅	5*							1				
G ₆	6*								1			
G ₇	7*									1		
G ₈	8*										1	

図2a 反応物結合行列

	1	9	10	2	11	12	3	4	5	6	7	8
	C	H	H	C	H	H	H	H	*	*	*	*
1C		1	1	1			1					
9H	1											
10H	1											
2C	1				1	1		1				
11H				1								
12H				1								
3H	1											
4H				1								
5*								1				
6*									1			
7*										1		
8*											1	

図2b 生成物結合行列

	1	9	10	2	11	12	3	4	5	6	7	8
	C	H	H	C	H	H	H	H	*	*	*	*
1C				-1			1					
9H												
10H												
2C	-1						1					
11H												
12H												
3H	1						-1					
4H				1			-1					
5*												
6*												
7*												
8*												

図2c 反応行列

ムである。そこで C-P 結合生成のために Arbusov 反応に関する規則が組み込まれており、実際に Phosphacarnegine の合成の最適経路の設計を行ない、それを合成実験と比較した結果を報告している³⁾。それによると 10 段におよぶ合成ルートの一部において副反応を避けたり、容易な反応で置き換える必要はあったが全体として有効であることを示した。

2-2. 行列方式

反応を化合物基準の表現にすることは処理の面では有利とはいえない。計算機に適した、より体系的な方法として Hendrickson は反応の行なわれる結合を切断してそれぞれを半反応 (half reaction) として捉え、化合物をそれでつなぐ方式で全体の体系化を試みている。その場合、化合物間の距離も明確となり、合成ルートの探索も機械的に行なえることになる⁴⁾。

また Ugi⁵⁾ や米田^{6),7)} は反応化合物と生成化合物をそれぞれ隣接行列で表わし、その差を反応を表わす行列とした。たとえば (1) 式の反応に対し反応系、生成系、反応をそれぞれ図 2 a, b, c のように表わす⁷⁾。この方式は演算が計算機向きであるだけでなく、例にあるように触媒を含めすべての関係を正確に記述できることが大きな特長である。また Ugi の EROS (Evaluation of Reactions for Organic Synthesis) においては行列の対角要素に非共有電子数、非対角要素には結合次数を記述する方式をとり、反応の行列要素の正負から、その結合が生成するのか切断されるのかも明らかに示せる。

2-3. CAMEO その他

LHASA が逆反応により合成経路をたどるのに対し、行列は前、後両方向が可能である。LHASA と同じ規則方式で前向き反応を処理するのが Jorgensen の開発した CAMEO (Computer Assisted Mechanistic Evaluation of Organic Reactions)⁸⁾ である。反応物質とその条件を与えると、反応規則を適用して生成物を出してくるシステムである。他の場合と同じようにグラフィック入出力機構をもち、入力データから官能基の特徴抽出を行ない、立体化学

や酸性度も識別する。有機ケイ素化合物を対象として酸、塩基触媒で求電子反応、求核反応などに適用される。CAMEO は化合物に即した従来形の記述形式をとっているため、half reaction や行列型に比して体系的ではないが、予測結果と実験との対比を行なっており満足すべき結果を得ている。

その他ニューヨーク大の Gelernter の SYNCHEM²⁰⁾ は評価に反応の難易のみでなく、収率、単価などを組み込み実用的である。また大型計算機上のシステムのほか、パーソナルコンピュータ版も開発している。トロント大の Bersohn^{10),11)} は、対話型でなく反応段数、収率を含め各種要求を制約条件として最適解を出すバッチシステムを開発している。

3. 反応のデータベース

反応設計が始められた頃は計算機の容量も小さく、またデータベースの概念も確立されていなかった。さらには少数のルールで、ある範囲がカバーできるとの考えから、データに重きが置かれていなかった。しかし複雑な反応を対象とすることや、反応規則の抽出に時間と高度な専門知識が必要なのがシステム実現の隘 (あい) 路となっていたことが認識されるようになってきた。一般的に反応データとしてオンラインで使えるものは ISI 社製の SDC 提供のものと Dercvent 社製の特許情報で同じ SDC が提供されているものなどがあるが、反応設計のシステムに結合されていないのでここでは除外することにする。

カリフォルニア大の Wipke¹²⁾ は LHASA の共同開発者でその方向の SECS (Simulation and Evaluation of Chemical Synthesis) を開発しているが、それとは別に REACS においては Organic Synthesis 全巻の反応を個別に入力し、反応を検索できるシステムになっている。しかも検索を図 3、図 4 に示すよう

図 3 SST プログラムによる合成ルート探索モデル

SM は出発物質, Eval は評価, m は分子, g はグラフを示す。

に化合物直接の照合ではなく抽象化されたグラフのレベルで人工知能的に行なうシステム SST (Super Structure Search) プログラムを開発し反応設計の“創造性”を高めることをねらっている。Cohen も反応のカタログ化を図り¹⁸⁾、部分構造の効率検索で成功した Dubois の DARC¹⁴⁾ も反応設計の知能化を目指している。同様な試みはわが国でも始められており松浦¹⁵⁾、筆者ら¹⁶⁾、原ら¹⁷⁾などの報告がある。

4. 反応設計の問題点

すでに述べたように反応設計における機能と処理能力は反応の表現に大きく依存する。とくに反応ルールの適用性も高めるには官能基とその周囲の結合とを、詳細にかつ、ある程度独立に記述できることが必要である。これは特許で問題となる化合物の総称表現 (generic representation) と共通である。Lynch¹⁸⁾,¹⁹⁾ はそのために化合物記述言語 GENSAL を開発し、特許における構造データの表現法を提案している。筆者らは化合物のグラフの特徴を利用して拡張 BCT 表現²⁰⁾,²¹⁾ によってこの問題に対応できることを報告している。また化学情報のうち重要な名称—化学構造の関係の両方向自動処理が知的反応設計には必須であるがそのためのシステム開発も進めている²²⁾。

5. むすび

計算機の高機能化と、データベースの整備を含めたソフトウェアの充実により、反応設計は実用化の時代に入り、事実、欧米では企業、研究機関でかなり使用されている。わが国においては今までは立ち遅れていたが、急速に普及のきざしが見えていることは喜ばしいことである。まだ未解決の課題も多いので本稿がこの問題に対する関心を高める契機になれば幸いである。

引用文献

- 1) E. J. Corey, W. T. Wipke : *Science*, **166**, 178 (1969)
- 2) D. A. Pensak, E. J. Corey : *Computer Assisted Organic Synthesis* (ACS Sym. Ser. 61) (1977)
- 3) C. Laurencio, L. Villien, G. Kaubman : *Tetrahedron*, **40**, 2721, 2740 (1984)
- 4) J. B. Hendrickson : *JCICS*, **19**, 129 (1979)
- 5) J. Blain, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi : *Tetrahedron*, **30**, 1845 (1974)
- 6) Y. Yoneda : ケモグラム, 丸善 (1972)
- 7) Y. Yoneda : *Bull. Chem. Soc. Japan*, **52**, 8 (1979)
- 8) C. E. Peishoff, W. L. Jorgensen : *J. Org. Chem.*, **50**, 1056 (1985)
- 9) K. K. Agarwal, D. L. Larsen, H. L. Gelern-ter : *Comput. & Chem.*, **2**, 75 (1978)
- 10) M. Bersohn, K. Mackey : *JCICS*, **19**, 137 (1979)
- 11) M. Bersohn : *Bull. Chem. Soc. Japan*, **45**, 1897 (1972)
- 12) W. T. Wipke, D. Rogers : *JCICS*, **24**, 74 (1984)
- 13) B. J. Cohen : *JCICS*, **22**, 195 (1982)
- 14) J. E. Dubois, Y. Sobel : *ibid.*, **25**, 326 (1985)
- 15) 松浦育敏 : 情報化学討論会, **6**, 56 (1983)
- 16) 内丸忠文, 佐藤 浩, 原 昭二 : 同上, **7**, 153 (1984)
- 17) 中山 堯, 藤原 譲 : 同上, **6**, 51 (1983)
- 18) J. M. Barnard, M. F. Lynch, S. M. Welford : *JCICS*, **24**, 66 (1984)
- 19) M. F. Lynch, J. M. Barnard, S. M. Welford : *ibid.*, **25**, 264 (1985)
- 20) T. Nakayama, Y. Fujiwara : *ibid.*, **23**, 80 (1983)
- 21) Y. Fujiwara, T. Nakayama, K. Hatada, A. Nishioka : *Data for Sci. Tech. (CODATA)*, **1983**, 150
- 22) 藤原 譲, 梶原康毅, 荒木啓介, 前田知子, 他 : 情報科学技術研究集会論文集, **21**, 75 (1984)