

## 知的医薬品設計エキスパートシステム IDDEX

岩井 憲一<sup>†</sup> 関谷 正明<sup>††,\*</sup> 野村 康雄<sup>††</sup>  
植田 育男<sup>†</sup> 溝口 理一郎<sup>†</sup>

一つの新薬を開発するには、薬理作用の改善や副作用の除去などの対策を考慮する必要から、多大の時間と労力を要する。そのため候補化合物は、できるだけ合理的かつ慎重に選択されねばならない。一般に、化学構造と薬理作用との相関関係は明らかではないので、新薬設計にはドメイン知識に加えて多くの経験的知識を必要とする。そこで従来より人工知能の手法を用いた医薬品設計システムの構築が望まれてきた。本稿では、スクラップ & ビルド法に基づく知的医薬品設計エキスパートシステム IDDEX について述べる。医薬品は他のドメインとは異なり、部品の概念が明示的に意識されていないのでドメイン知識が部品を中心に整理されていない。そこで本研究では、与えられた化合物（リード化合物）をいくつかの部分構造（部品）に分解し、その部品を置換可能な別の部品と置換した後再結合するというスクラップ & ビルド法を導入することによって、医薬品設計を既存の部品の組合せとして定式化した。また IDDEX の利用を通じて部品のデータベース化を行い、新薬設計に必要な知識の獲得を行った。その結果、部品に内在する知識を明示的に取り扱うことが可能となり、従来の計算機支援による医薬品設計より高いレベルの設計を合理的に行うことが可能となった。

### IDDEX: Intelligent Drug Design EXpert based on Scrap & Build Method

KENICHI IWAI,<sup>†</sup> MASA AKI SEKITANI,<sup>††,\*</sup> YASUO NOMURA,<sup>††</sup>  
IKUO UEDA<sup>†</sup> and RICHIRO MIZOGUCHI<sup>†</sup>

Designing a new drug requires a lot of time and effort, since developers have to make up a number of counterplans for improvement of biological properties, elimination of side effects and so on. Therefore, candidate compounds should be selected as carefully as possible to avoid unnecessary examination which costs a lot of money. Because a correlation between a chemical structure and a drug activity is not fully known, creating a new drug needs many heuristics in addition to domain theory. Artificial intelligence techniques enables us to develop such a computer aided drug design system that supports domain experts. In this paper, we propose an expert system IDDEX: Intelligent Drug Design EXpert for drug design based on scrap & build method. So far, domain knowledge of drug is not based on parts. And we formulate drug design as combination known parts by scrap & build method that decomposes a given lead compound using a suitable sub-structure compound and rebuilds a new candidate compound by substituting the resulting sub-structure of the lead compound with appropriate parts. Through use of IDDEX, the drug designer can make a data-base for parts, and acquire knowledge for drug design. As a result, we can clearly process potential knowledge in parts, and can rationally design a new drug.

#### 1. ま え が き

一つの新薬を開発するには、以前より優れた主作用の獲得・副作用の除去・未知の疾病に対する対策等の

必要性から、極限的な性質や精密な仕様を満たすことが要求されている。一般に、医薬品における化学構造と薬理作用との相関関係、いわゆる構造活性相関は明らかではない。そのため、現在一つの薬を作るために平均して約 20,000 個もの化合物をスクリーニングする必要があるといわれており<sup>1)</sup>、医薬品設計における専門家の負担は増大する傾向にある。

このことより、試作される候補化合物をできるだけ精選し、優れた生理活性を示す化合物をできるだけ高

<sup>†</sup> 大阪大学産業科学研究所  
The Institute of Scientific and Industrial Research,  
Osaka University

<sup>††</sup> 関西大学工学部  
Faculty of Engineering, Kansai University

\* 現在、シャープ(株)  
Presently with Sharp Co., Ltd.

い確率で獲得しうる合理的な手法が必要となるが、そのためには、現在までに得られた経験則を蓄積し、これを積極的に活用することが望ましいと考えられる。このような手法、つまり人工知能の手法を用いた計算機支援による医薬品設計が、従来から切望されている。

われわれは、これまで材料設計に対して人工知能の手法を用いた設計支援の研究を行ってきた<sup>1)-3), 9)-11)</sup>。本稿ではドメインを医薬品に限定し、医薬品設計に対して計算機支援を実現するために導入した新しい概念であるスクラップ&ビルドの概念に基づいた知的医薬品設計エキスパートシステム IDDEX (Intelligent Drug Design EXpert) の概要について述べる。

## 2. 基本概念

### 2.1 ドラッグデザイン

ドラッグデザインを行うには、さまざまなアプローチが提案されているが、通常は新しい薬を設計すること、すなわち創薬における何らかのきっかけとなるべき原材料を出発点とし、これに対して構造修飾を行うことが多い。この出発点となる原材料の化合物は、リード化合物と呼ばれる。このことから、医薬品の設計をリード化合物について着目すると、次の三つのクラスに大別される。

- (1) リード化合物の生成 (lead generation)
- (2) リード化合物の展開 (lead evolution)
- (3) リード化合物の最適化 (lead optimization)

(1)の生成は、新しくリード化合物を見いだす場合の設計であり、(2)の展開は、与えられたリード化合物に対して構造修飾や置換を行うことにより、新しい高次のリード化合物を作成する場合の設計であり、(3)の最適化は、物性パラメータ等の操作や系統的修飾により活性の優れた関連化合物に導く場合の設計である。

また一般に設計問題は、設計すべきシステムとそれを構成する部品に着目すると、次の三つのクラスに分類できる。

- (1) システムの構造と部品が未知のクラス
- (2) システムの構造が未知で部品が既知のクラス
- (3) システムの構造と部品が既知で部品のパラメータが未決定のクラス

これに照らし合わせてドラッグデザインを考察すれば、同じ番号どうしが対応しており、この順に創造性が必要とされる魅力的な研究対象となっていることが

わかる。これまでコンピュータを利用した設計支援システムに関する研究は盛んに行われてきたが、そのほとんどは(3)のタイプのものであった。これは、(3)の最適化においては構造活性相関である程度解明されており、すでに QSAR (Quantitative Structure-Activity Relationship: 定量的構造活性相関) と呼ばれる方法論が確立されているからである。QSAR は化学構造と生物活性の相関関係を数式で表現しようとするものである。ところが、QSAR が有効なのはあくまで(3)の最適化の場合であり、薬効の高い画期的な医薬品を開発するにはリード化合物の展開以上の上流の設計行為が不可欠となることはいうまでもないが、(1)の生成のように何の手掛かりもなくいきなり新しいリード化合物を見いだすことは、複雑かつ広範囲にわたる知識と人間固有の類推やひらめきが必要となり、現在の情報処理技術ではその実現が困難である。このような考察の結果、本研究ではリード化合物の展開の支援を最終目的としている。

### 2.2 基本となる概念

ドラッグデザインは従来から実験先導型であり、設計に必要な知識は実験を通じて得られた研究者の経験や knowhow が中心となっている。しかしこれらの知識を計算機上で用いるのは非常に困難であった。その理由としては、他の分野で実用されている設計支援システムとは異なり、明確な設計知識の抽出と記述が困難であること、そしてドメインにおいて部品の概念が明示的に意識されていないため、ドメイン知識が部品を中心に整理されてはいないことなどが挙げられる。そこで本研究ではこの2点を解決するために以下の概念を導入した。

#### (1) スクラップ&ビルド法

まず最初の概念は“スクラップ&ビルド法”である。この概念は設計を行う際に仕様と原材料であるリード化合物が与えられると、このリード化合物の部分構造をもつ化合物によって分解を行い(スクラップ)、この分解によって得られた化合物に対して置換や構造修飾等を行う(ビルド)ことにより、新薬の設計を行っていくというものである。

従来のドラッグデザインは、主に既存の医薬品を出発点とし、物性パラメータの最適化や原子レベルの組み替えによって行われていた。この結果、原子間の組合せ爆発が生じたり、また同じ原子でも結合の部位によって作用が異なるので系統的な設計が困難となっていた。そこでこのような問題を解決するために、設計

に分解の概念を導入した。

設計の最大の共通概念として“設計は部品の組合せである”というものがあるが、そこでは“部品”という概念が重要な役割を担っている。薬の構成は骨格・ブリッジ・アミン基の三つに分けて考えることができるが、そのいずれもが部品と考えることができ、薬効の発現に関してそれぞれ独自の役割を担っている。図1に実際の医薬品を骨格・ブリッジ・アミン基の三つに分割した例を示す。

部品のグレインサイズを原子からある程度の大きさをもつ部品に置き換えることにより、その基本構造の数は増大するが逆に接続可能性のある部品の組み合わせ数は激減し、全体の処理時間の大幅な削減が可能となる。また部品の置換を行うことによって各部品に何らかの等価性や意味づけを行うことが可能となるので、構造活性相関の手がかりを得るきっかけになり、計算機支援に何らかの方向性をもたせることができる。つまり各部品に意味づけを行った上で、ある医薬品の主作用を示す部品（以下、主作用団もしくは骨格 (skeleton) と呼ぶ）や、主作用を増強する作用を示す部品（以下、副作用団と呼ぶ）として蓄えておけば、最も効率のよい設計を行うことができると考えられる。

この概念を本研究に導入することによって、医薬品設計のプロセスを分解と生成の二つに定式化することができる。

また生成を効率良く行うには、分解によって得られた部品を系統的に整理する必要がある。本研究では、その判断基準として isosterism（等価性）の概念を導入している。isosterism は主に次の二つに分かれる。

- (a) bio-isosterism（生物学的等価性）
- (b) chemical-isosterism（化学的等価性）

(a) は、過去の設計においてある化合物の一部分を別の部品に置き換えても、薬理作用に変化が見られない例が一つでもあれば、この二つの部品は生物学的に等価であるとみなして一つのカテゴリーにまとめる概念である。図2に bio-isosterism の概念に基づいて等

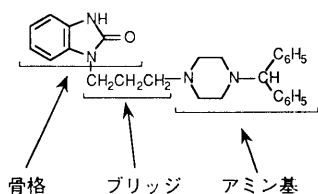


図1 医薬品の構成

Fig. 1 The structure of a drug.

価とみなされる化学構造の例を示す。

(b) は、例えばその価電子数が等しいというように、化学的性質が同じ化合物を化学的に等価であるとみなす概念である。

以上のような二つの観点から等価な化学構造どうして構造の置換を行い再合成を図っていく。

## (2) CBR (Case-Based Reasoning: 事例ベース推論)

一般に専門家が何か新しい設計を行うときには、以前に設計した類似設計例に基づいてそれを適宜修正して行う場合が多い。そこで専門家が行う設計過程を“類似設計”としてとらえ、これをシステム上に実現することによって専門家への設計支援を行うことを試みる。そこでわれわれが採用するもう一つの方法が“CBR (Case-Based Reasoning: 事例ベース推論)”である。

CBR は過去の事例を事例ベース上に蓄積し、これを陽に利用して、与えられた問題と類似した事例を検索してその問題との差異を評価し、その差を埋めるべく修正した後また評価・修正を繰り返すという検索・評価・修正の3過程の枠組みを用いることにより問題解決を行うという手法である<sup>12)</sup>。この手法は特に設計問題のような合成型問題に有効である。ルールのみを用いた推論では、最適解が定まりにくく、組合せ爆発する可能性が大きい問題では、事例ベース推論を利用することによりそれを防ぐことができる。

以上の概念を導入することによって、計算機支援が可能となるようなドメイン知識の再構成と設計知識の抽出困難性を補うことが可能となる。

## 3. システムの概要

### 3.1 システムの設計思想

筆者らは、これまで人工知能技術を駆使した巧妙な設計エキスパートシステムの設計を目指して研究を進めてきた。そのアプローチはトップダウンな方法といえることができるが、他の多くの人工知能研究者も同様のアプローチを採ってきたといえる。一般に、優れたエキスパートシステムの開発には、シーズとニーズのバランスの取れた設計および取り扱うデータや知識の

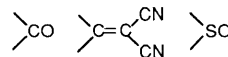


図2 bio-isosterism の観点で等価な化学構造の例  
Fig. 2 Examples of the equivalent chemical structure on bio-isosterism.

整理が不可欠となることは明らかである。それにもかかわらず、これまで実社会で開発されてきたエキスパートシステムの多くがプロトタイプで終わっている原因に、ニーズの正確な把握が行われていないことと、専門家がシステム構築時に本当に必要な知識を明示的に記述するのが困難であることが挙げられる。本研究では、当研究所の医薬品設計の専門家との共同研究の実施によって、人工知能技術で実現できる機能やシステム構成を提案しながら、専門家が本当に必要としているシステムを構築することを念頭に置いた設計支援システムの構成を検討してきた。

何度が専門家と議論することによって、次のことが明らかになった<sup>1)</sup>。

- (1) 自分の代わりに設計を代行してくれるシステムではなく、優れた助手として機能するシステムを望んでいること。

新薬設計の支援システムを構築する際に注意すべきことの一つに、知識工学者は新薬自動設計システムを構築しがちであることが挙げられる。しかしこれは現状では非常に困難であるだけでなく、新薬を設計するのはあくまでも専門家自身であり、システムは専門家の発想や設計を環境面から支援するという形が新薬設計支援の最善の形である。

- (2) 専門家の経験的知識の本質は、設計における部品となる医薬品の部分構造の認識にあること。すなわち、組合せ爆発を防ぐほど大きなサイズで、かつ多様な設計を行うことができる範囲の小さな部品単位の設定が不可欠であること。

2章でも述べたが、設計問題に対する計算機支援を行うにはその環境整備が必要である。専門家と従来のシステムの大きな違いの一つは、専門家は部品を見ただけでその部品に内在する知識や組替えの指針が認識できることである。そのためこれを計算機上で実現するためには、まず全医薬品を分解して膨大な数の部品を作成し、これを体系的に整理する必要がある。これによりドラッグデザインを系統的に行うことが可能となる。

- (3) 設計時における検索を的確に行うために、その部品への適切なインデックスの付与、およびそのデータベースの整備が重要であること。

部品を利用してドラッグデザインを行うには、部品に内在する知識を利用できる形に定式化する必要がある。そのため本研究では、部品群をデータベース化し、部品に内在する知識を整理してインデックスの形

で部品に付与していく。

- (4) 過去の医薬品設計の成功例などの解析を通じて、設計者独自の部品データベースをいつでも容易に更新できる環境が必要であること。

新薬設計を行うには、設計中に随時新しい知識や部品を必要とする。そのため、システムは常に専門家の要求する部品や知識を提示したり獲得する環境が不可欠となるので、このことにいつでも対処でき、専門家にとって利用しやすいシステム構成が必要である。

- (5) 医薬品の分子構造を表示する機能を含めた優れたインタフェースが不可欠であること。

専門家は常に化合物の薬理作用を考慮しながらその構造を修正していくので、その化学構造とその属性および置換に用いられた部品等を専門家に提示する必要がある。

- (6) 新薬を設計する際に本当に必要な知識を専門家が明示的に記述することは困難であるので、利用を通じて知識獲得のためのデータと事例を収集し、後で筆者らが知識ベース化する方式が有効であること。

部品群がある程度充実してくると、今度はその部品の利用の仕方に関する知識が必要になるが、特に従来の設計知識に加えて今までにない新しい知識を随時取り入れる必要がある。一般に新知識には、その時点ではうまく説明できない“主観”が含まれている場合が多いため、従来の知識獲得手法のようにただ専門家に直接必要な知識を聞きだすだけでは知識獲得ができない。そこで本研究では、その前段階として本システムの利用時の各過程において専門家が用いたデータや事例およびその説明や理由を専門家に記述してもらう手法を採用した。そして集められた文章を筆者らで解析してキーワードを抽出し、このキーワードを基に以後の知識獲得を行っていけば、専門家が示せなかった知識が得られると考えられる。

以上のような検討の結果を踏まえて、われわれは従来にない新しい方法に基づく知的医薬品設計エキスパートシステム IDDEX (Intelligent Drug Design EXpert) を構築した。以下では IDDEX の概要について述べる。

### 3.2 システムの構成と動作の概要

IDDEX の構成を図 3 に示す。IDDEX は Scrap & Build モジュール (以下, SBM) と部品データベース (以下, PDB), リード化合物データベース (以下, LDB), CBR の概念に基づいた推論を行う推論エンジ

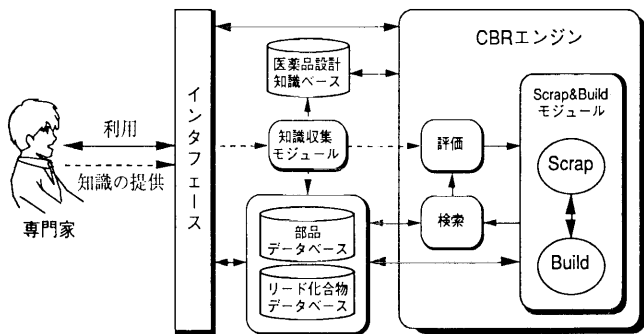


図 3 IDDEX の構成図  
Fig. 3 Architecture of IDDEX.

ン (以下, CBR エンジン), 医薬品設計知識ベース (以下, DDKB), そして知識収集モジュール (以下, KAM) から構成され, ユーザとの間にインタフェースを介している. IDDEX の位置づけとしてはユーザを専門家に限定しており, 専門家に対してインタラクティブな新薬設計支援を目指している.

IDDEX の入出力を以下に示す.

入力: リード化合物の構造式, 設計仕様, 骨格の構造式

出力: 候補化合物群

入力には新薬の原材料となるリード化合物および骨格の構造式と要求仕様が用いられる. 仕様には設計のゴールが記述されている. IDDEX で扱う設計のゴールは薬理作用の改善であり, それを実現する新薬の構造式の生成である. 薬理作用の例を図 4 に示す.

薬理作用は主に主作用・副作用・毒性除去・作用持続・吸収改善・安定性の六つの項目からなる. 設計の際に専門家が用いる知識もこれらの 6 項目に沿っている.

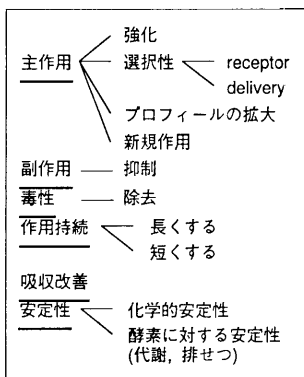


図 4 薬理作用の例  
Fig. 4 An example of drug activities.

出力は新薬候補の化合物群であるが, 実際には要求仕様を満足するまでこの候補化合物を次の新たなリード化合物に利用するという過程を繰り返している.

IDDEX が行う動作の流れを図 5 に示す. IDDEX の動作には設計支援モードと部品充実モードの 2 モードがある. 部品充実モードは設計支援の前段階として行われるが, 詳細は CBR エンジンの節で述べる. 設計支援モードについては以下のとおりである. まず専門家が IDDEX に要求仕様を入力すると, 該当する

リード化合物の候補群を IDDEX が LDB から検索し, 専門家に提示する. この中で専門家が適切と思うものがあればそれを選び出すが, 適切なものがなければ専門家自らがリード化合物の構造式を作成し, これを IDDEX への入力とすることができる. 次に専門家

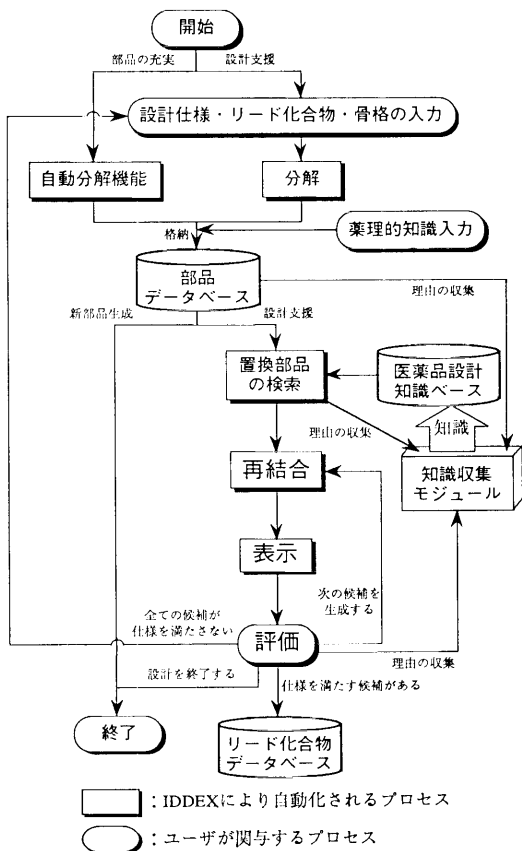


図 5 IDDEX の動作の流れ図  
Fig. 5 Flowchart of IDDEX.

はリード化合物の構造を見て、骨格と思われる部分をリード化合物の構造から選び出し、それを骨格としてIDDEXに入力する。

IDDEXは入力としてリード化合物および骨格の構造式と要求仕様が与えられると、図6に示すSBMのScrap Moduleを用いてリード化合物を分解する。

SBMの構成は図6に示すようにScrap ModuleとBuild Moduleからなる。Scrap Moduleへの入力はリード化合物と骨格であり、出力は分解して得られる部品群である。またBuild Moduleへの入力はこの部品群を置換して得られた部品群と骨格であり、出力は置換後の部品群と骨格を再結合して得られた新薬候補化合物である。SBMは主に次の場合に利用される。

- (1) 医薬品設計支援
- (2) 部品データベースの拡充

Scrap Moduleにおいてリード化合物を骨格で分解するが、骨格は専門家によって示されたリード化合物の部分構造であるので、IDDEXにとってはその属性が不明な未知の部品である。新薬設計を行うのであればこの骨格に内在する知識(属性)が不可欠であり、この属性に基づいて設計が行われるので、KAMが専門家にインタビューする必要がある。また今後の研究材料として、なぜこの骨格構造を骨格としたのかについてのインタビューも併せて行っている。そして元の部品を検索された部品で置換した後Build Moduleを用いて再結合し、これによって得られた新しいいくつかの化合物を新薬候補として専門家に提示する。専門家は、満足できる結果が得られるまでこの操作を繰り返す。

ただし構造活性相関が不明確なため、得られた解候補が実際に有効であるとは限らないので、IDDEXの利用により候補化合物群が生成されると、専門家がこの中で薬理作用がある可能性が高いと判断したものを

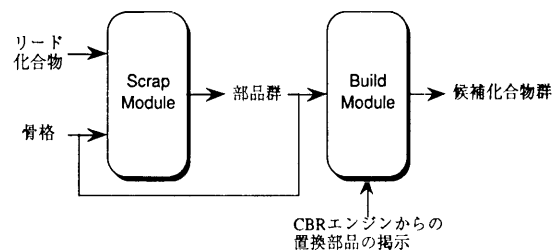


図6 Scrap&Build モジュールの流れ図  
Fig. 6 Flowchart of Scrap&Build module.

実際に作成し、薬効判定試験を行う必要がある。この際にもKAMがインタビューすることにより、常に評価結果の情報を取り込み、知識の質の向上を図っている。

次節では、本システムの各モジュールの機能について述べる。

### 3.3 PDBとLDB

一般に、専門家は化合物を見ただけで、そこに内在する豊富な知識を取り出して扱うことができる。しかし計算機にはそのような振舞いは困難であるので、そのような知識はあらかじめ明示的に表現しておく必要がある。このような観点に基づいて検討されてきたのが、部品データベース(PDB: Parts Data Base)とリード化合物データベース(LDB: Lead compound Data Base)である。PDBは、医薬品設計に必要な基本的部品と、SBMによって得られた部品をデータベース化したものであり、LDBはリード化合物やIDDEXの利用によって作成された解候補の中で有用なものをデータベース化したものである。データ構造は、部品データ・リード化合物データ共々化学構造の接続情報とそれに付随するインデックスからなるので、本稿では部品データについて説明する。

#### (1) 部品データ

部品データは、化合物の構造式を、構成する各原子間の接続情報(図7参照)とそれに付随する各種インデックス(図8参照)によって表現する。部品そのものは、その医薬品構成の観点から大別すると、2章で述べたように、何らかの薬理効果を示す主作用団(骨格)と、その主作用団の薬理効果を増強、抑制または持続等を行う副作用団に分かれる。また作成過程の違いによって、例えば官能基のように主に化学的に重要な部品やドラッグデザインに頻繁に用いられる部品である基本部品データと、ユーザがIDDEXの利用時などで必要と認めて独自に登録したユーザ定義部品データに分類できる。前者は薬理的性質は不明確であるが化学的性質はよくわかっているので汎用的に多種の設計に用いられ、後者は化学的性質や薬理的性質はよくわかっているが、特定の仕様設計にしか有効でないという特徴がある。なお現在格納されている基本部品データの数はおよそ100種類であり、必要な量の約50%をカバーしている。

既存部品による設計を行う際に専門家が参照するものとして過去の事例がある。事例は設計された薬の機能と活性の対、あるいは設計過程全体であるというこ

とができるが、専門家の利用法を分析した結果、事例を直接用いているのではなく、設計事例において重要な役割を演じた部品の置換、あるいは部品そのものによって還元された形で用いていることが明らかとなった。

このことより IDDEX の事例は知識を付加した部品となり、PDB は設計を繰り返すことによってより有用なデータベースへと成長する。そして CBR エンジンには、部品に対してそのインデックスの情報をもとに検索・修正・評価という CBR 本来の動作を繰り返すことによって部品検索を行い、ユーザに適切な部品を示すことができる。

リード化合物はいうまでもなく過去の事例であるので、このことから IDDEX の事例ベースは PDB と LDB が担っているといえる。

新薬創製のためにはまず設計の前に有用な事例の整備と拡充が不可欠である。IDDEX の利用によって従来にはない有用な部品やリード化合物が得られた場合にはそれを PDB や LDB に登録しておく必要がある。

新しい部品データは、登録の際にデータが重複していないかを確認しておけば、従来のような構造によってデータを検索する必要がなく、部品データに付随するインデックスによる検索となるので、PDB は通常の関係データベースにまで縮約され、検索に要する時間を大幅に削減することが可能となる。

また、従来の多くの医薬品設計ではこの接続情報に準ずるもののみをデータ構造として採用していたが、これでは例えば互いに鏡像の関係にあるような異性体である光学異性体に対する処理が困難であった。同じ接続関係を持つために従来の接続関係のみのデータ構造では見分けがつきにくく、中には特性が著しく異なる化合物も存在している。

本研究ではこのような問題点を解決するために、化学構造だけでは判別がつきにくい化合物の情報を、インデックスの形で部品に付加しておく方法を採用している。これにより、より高度で複雑な問題を扱うことができる。

そこで、以下では部品データの接続情報とインデックスについて述べる。

### (2) 部品データの接続情報

図 7 に部品データの接続情報を示す。部品データの書式は ChemDraw Plus<sup>16)</sup>

の Connection Table 形式を採用している。その詳細は、上から順に化合物名・原子数および原子間の接続数、そして各原子の X 座標・Y 座標・Z 座標と原子名、それからこの各原子に対して上から順に数えたときに、何番目と何番目の原子が何重結合でどのような状態で結合しているかを表している。

いま図 7 ではサリチル酸を例に挙げているが、図中のような化学構造式を入力すると以下のような部品データに変換される。

### (3) 部品データのインデックス (薬理的知識)

インデックスの項目を図 8 に示す。大分類とは薬理的に同じカテゴリーに分類される部品を含み、既存薬の最も基本的な分類になっている。中分類とは薬理

化合物名	結合数			X座標	Y座標	Z座標	原子名 (部品)
サリチル酸	8	8					
原子数	-0.4361	0.5000	0.0000	C			
	-1.3028	0.0000	0.0000	C			
	-1.3028	-1.0000	0.0000	C			
	-0.4361	-1.5000	0.0000	C			
	0.4305	-1.0000	0.0000	C			
	0.4305	0.0000	0.0000	C			
	-0.4361	1.5000	0.0000	OH			
	1.3028	0.5000	0.0000	COOH			

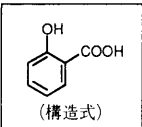
  

1	2	2	0
2	3	1	1
3	4	2	0
4	5	1	1
5	6	2	0
1	6	1	1
1	7	1	1
6	8	1	1

1番の原子と2番の原子が結合している

2重結合している

結合の状態



(構造式)

図 7 部品データの接続情報

Fig. 7 The connecting information of parts data.

Serial No.	143	ファイル名	C23H19N5O5Cl
設計日時	93年 8月 4日	置換可能ファイル名	p86
	12時 43分	評価結果	鎮痛、頭痛、麻酔作用
大分類	4	親のリード化合物	a
中分類	3	親の骨格	b
小分類	2	要求仕様	主作用一作用の増強
構造分類 1	C		
構造分類 2	A		
化学式	C23H19N5O5Cl		
正式名称	C23H19N5O5Cl		
慣用名	C23H19N5O5Cl		
分子式	C12COOHHSN2C4		
分子量	219		

図 8 部品データのインデックスの例

Fig. 8 An example of the parts data index.

学的に同じ作用プロフィールを示す薬物や、薬物の作用部位が類似している薬物の集合である。小分類とは、中分類に分類された薬物を、さらに薬理学的に類似している薬物群や投与形態が類似した薬物を一括したものである。

次に構造の面から見ると、構造分類1とは小分類された薬物を化学構造の特性や構造的な類似性をもとにグループ分けしたものである。構造分類2とは先の構造分類1でグループ分けされた薬物を個別の薬物の化学名、一般名または商品名とその構造で分類したものである。既存の薬物に関しては、大分類で20種類、中分類で10種類、小分類で多くて12種類程度で分類できることがわかっている。

上述の既存の薬品における類似性の判断は専門家固有のものであり、インデックス自体に専門家の know-how が内在するということができる。

### 3.4 CBR エンジン

CBR エンジンは CBR の概念に基づいた推論を行う推論エンジンであり、主に以下の二つの機能を持っている。

- (1) 仕様に適切な置換可能部品やリード化合物の検索
- (2) 新部品の PDB への格納と新薬候補の LDB への格納

本エンジンは CBR の枠組みに基づいているので、推論過程は検索・評価・修正（操作）の3過程からなり、これを実現する三つのモジュールからなる。CBR エンジンはこれらの操作を繰り返し実行することにより、問題解決を行っている。これらの推論過程の中で、修正は SBM が行っている。部品置換の際に過去の事例が必要な場合や要求仕様からリード化合物を検索する場合は、検索モジュールを用いてそれぞれ PDB と LDB から検索を行っている。検索モジュールは、通常のデータベースにおけるテキスト検索機能だけでなく、SBM の Scrap 機能を利用した自動構造検索機能をもつ。これは例えば前節で述べたように分解によって得られた部品を PDB に格納する際にその部品が新規であるかどうかを判断する場合や、また特定の部品を部分構造にもつリード化合物を LDB から検索する場合に用いられる。

また 3.1 節でも述べたように、効率の良い設計を行うには有効な部品群、つまり PDB の充実が不可欠である。CBR エンジンにはこのために、LDB 内のリード化合物の PDB 内の全部品で分解する自動分解機能

をもつ。評価モジュールは、修正によって得られた候補を評価するものであるが、すでに述べたように医薬品設計においては構造活性相関が明らかではないことから現段階では評価知識の獲得が困難なため、専門家に評価を委ねている。

CBR エンジンが取り扱う知識は医薬品の薬理作用に関する知識と、構造修飾に関する知識の2種類に分かれるが、筆者らは便宜上それぞれを薬理的知識・医薬品設計知識と呼んでいる<sup>11)</sup>。本研究では、薬理的知識は部品データやリード化合物データに付随するインデックスの形で、医薬品設計知識は次節で述べる DDKB に蓄積されている。

### 3.5 DDKB と KAM

医薬品設計知識ベース (DDKB: Drug Design Knowledge Base) は、部品データを取り扱う知識を知識ベース化したものである。

医薬品設計知識は構造修飾の規範となる知識であり、一般にその形態はルールと制約と数式に分類される。ルールは専門家の経験則であり、制約は置換基の電子効果のような化学的性質を指し、数式は QSAR 法等において定量的な最適化に用いられる。しかし本研究ではリード化合物の最適化は行わないので、ここでは数式は考慮しない。医薬品設計知識は、ルール・制約の形で整理中であり、現時点で約 30 個ある。

また新薬設計には、随時新しい知識が必要となる。そこで筆者らは専門家が IDDEX を用いて設計を行う過程でその操作の理由を獲得する知識収集モジュール (KAM: Knowledge Acquisition Module) を構築中である。

KAM によって獲得された情報は次回以降の設計方針やインタビューに活かすために筆者らで解析して知識に変換しておく。このモジュールにより新薬設計に必要な知識を得ることが可能となると考えられる。しかし、KAM およびそこからの知識獲得は現時点では検討が十分ではなく、今後の課題である。

## 4. 設計支援

本章では、IDDEX が行う設計支援の方法を実際の例を用いて説明し、次章でインプリメンテーションについて述べる。

### (1) 仕様選択とリード化合物の作成

まずユーザは、IDDEX に対して、これから行うべき設計に対する要求仕様を薬理的知識インタフェースから入力する (図 9 参照)。



そしてこの仕様に該当するリード化合物を LDB から検索し、ない場合はその構造式をリード化合物作成 Window 上でマウスを用いて作成し、またその中で重要と思われる部品構造をマウスで囲んで指定し、その部分構造を骨格作成 Window 上にペーストすることによって骨格を作成する。図9では専門家がピフェニール化合物を作成しその骨格としてカルボン酸となる部位を選択している。

### (2) 骨格によるリード化合物の分解

そしてこの二つの化合物を SBM の Scrap 機能を用いて分解を行う。本節の例では、分解の結果三つの化合物が部品として生成される。分解によって得られた新部品は、図10の化合物表示 Window を用いてユーザに提示される。

なお、IDDEX が骨格を用いてリード化合物を分解する場合には、ドラッグデザインを行う場合と PDB を拡充する場合がある。PDB 拡充の場合は、3.4 節でも述べたように CBR エンジンの自動分解機能を利用して次々に新しい部品の獲得を行う。

### (3) 骨格の属性入力

IDDEX によってリード化合物が骨格を用いて分解されると、専門家は図9の薬理的知識インタフェースを用いて骨格の薬理的知識を入力する。分解によって得られた新部品の中でユーザが有用と認めたものは PDB に格納することができる。

新部品は、格納の前にあらかじめ IDDEX によって部品データベース内の部品と同一のものでないかチェックされ、そうでないものだけ、インデックスの形で

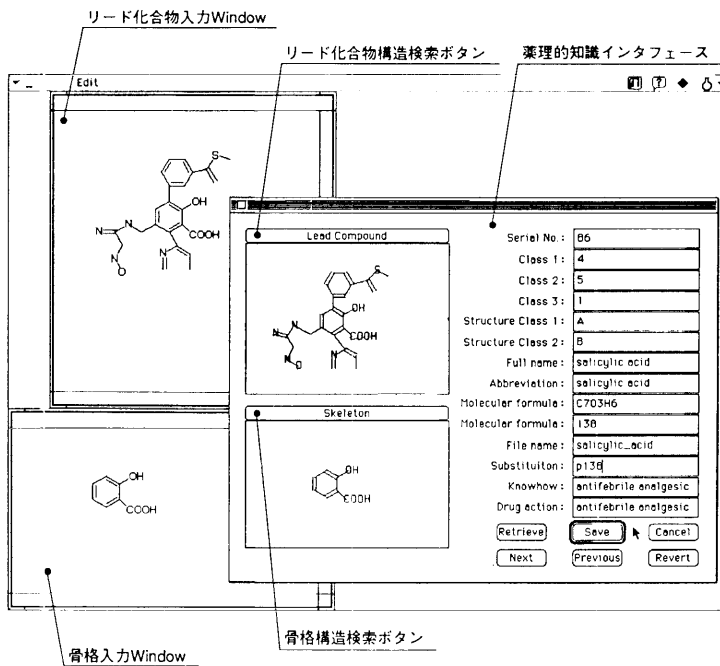


図9 薬理的知識インタフェース  
Fig. 9 Drug action knowledge interface.

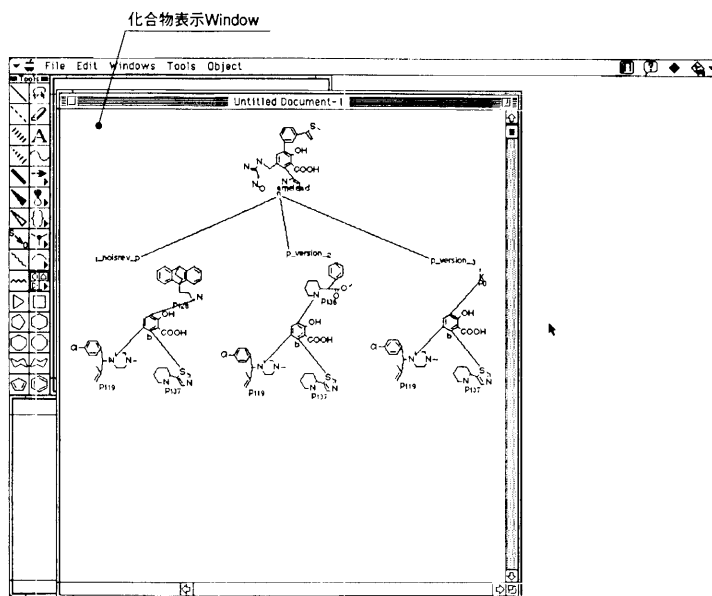


図10 化合物表示 Window  
Fig. 10 Compound display window.

薬理的知識の付加を行った後、格納する。その際にも図9の薬理的知識インタフェースを用いる。

#### (4) 部品の置換と新薬候補の提示

骨格の薬理的知識が与えられると, IDDEX はこの知識を基に置換部品を検索し, 元の部品との置換を行う. IDDEX における部品の置換の仕方には 2 種類ある. 一つは検索された全部品を元の部品と置換して再結合する方法である. このとき用いられた知識が専門家に提示される. そしてもう一つは, 部品検索時に専門家に部品群を提示し, その中から特定の部品を専門家に選ばせる方法である. この場合は専門家に選んだ部品についてインタビューする.

選択された候補部品群は再結合され, 専門家に置換の履歴を tree 表示の形で化合物表示 Window を用いて提示する. 図 10 に, 分解で得られた部品に対して maprotiline・methyphenidate・hydroxyzine・pipradrol・potassium などが置換部品として選択されこれを用いて再結合し, 三つの新薬候補が生成された例を示す. 専門家はその履歴を見ながら必要であれば履歴の修正を行ったり, その中から次のリード化合物を取りだして新しい設計に利用していく. また設計中に得られた部品や化合物の中で専門家が有効と認めたものは, PDB や LDB に格納される.

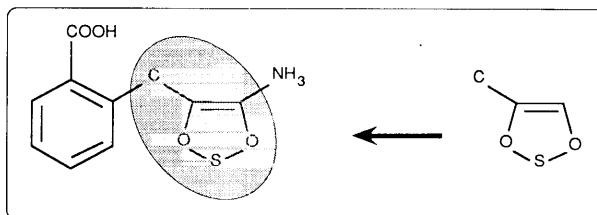
以上の操作を繰り返し行うことにより, ドラッグデザインを推し進めていく. 専門家はこの結果を基に, 実際に試作し薬効判定試験を行う. その評価結果を知識収集インターフェースに専門家が入力することにより次回以降の設計ではこの知識が有効となる.

### 5. インプリメンテーション

以上のようなシステム設計に基づいて, IDDEX のインプリメントを行ってきた. 本章では, 現在までのインプリメンテーションについて述べる.

#### 5.1 プログラミング環境

IDDEX は, Apple Computer, Inc. の Macintosh Centris 650 上で, SYMANTEC Corporation の THINK C を用いてインプリメントが行われてきた. 現在までに, まず IDDEX の中心的役割を果たす SBM と, それを管理しかつ置換部品を選択する推論エンジンである CBR エンジン, 部品とリード化合物をそれぞれ蓄積したデータベースである PDB と LDB, そして分解された部品や新たに生成された新薬候補化合物を表示する表示モジュールを構築した. これにより, 有効な部品の獲得および候補化合物群の作



左の化合物内に, 右の化合物が含まれるか調べる

図 11 サブグラフ同定問題の例

Fig. 11 An example of subgraph identification problem.

成が可能となった. また部品や骨格およびリード化合物の薬理的知識や医療品設計知識の収集を行う KAM をインプリメント中である. なおデータ構造の書式と, インターフェースの一部である化合物表示 Window の一部とリード化合物作成 Window および骨格作成 Window に Cambridge Scientific Computing の ChemDraw Plus を用いて実現した.

#### 5.2 サブグラフ同定問題

IDDEX の設計支援には, 化学構造の分解が不可欠であるが, これを実現するためには, リード化合物のどの部分構造が骨格か同定する必要がある. われわれはこれをグラフ理論におけるサブグラフ同定問題として捉えている. サブグラフ同定問題の例を図 11 に示す.

一般にサブグラフ同定問題は NP 完全問題であるので, われわれは, 原子名や原子間の接続情報等に注目して探索空間の絞り込みを行うことにより, 問題解決にかかる時間の短縮を図っている.

本研究では, グラフ理論のラベル付き無向グラフを構造式のモデルとして採用し, あるグラフの中に別のグラフが含まれるかどうかを判定するサブグラフ同定問題に置き換えて問題解決を行った.

### 6. 他の研究との比較

IDDEX と今まで開発された医薬品設計支援システム\*との比較を行う. 従来より医薬品設計支援の名のもとにさまざまな支援システムが試みられてきた. 例えば分子式から構造式を推定する佐々木の構造決定システム CHEMICS<sup>14)</sup> や化学エキスパートシステム構築のために化学構造式を表現し操作のための組込み述語を備えた支援ツールである阿久津らの化学エキスパート・システム構築ツール CHEMILOG<sup>15)</sup> のように設計環境に対する支援システムや MDL 社の

\* 化学構造式作成エディタや化学エキスパートシステムも含む.

MACCS (Molecular ACCess System)<sup>7)</sup> や住友化学と日本電気が開発中の ACACS (Advanced Computer Aided Chemistry System)<sup>7)</sup> などが挙げられるが、どれもリード化合物の最適化について支援したものが多く、専門家は、このリード化合物の最適化に対する計算機支援は可能であり事実上ほぼその最適化範囲を網羅していると評価しており、医薬品設計における最適化に関する研究は概ね完成しているといえる。しかし、生成や展開のような新薬設計を行うにはまだまだ研究の余地が残されている。

IDDEX が従来のシステムと異なる点は、まず医薬品設計のレベルをリード化合物の展開に置いたことが挙げられる。IDDEX と他のシステムの関係としては IDDEX の利用によって得られた出力結果である候補化合物を専門家がその生物活性を評価して活性があると確認できたものが従来の最適化システムの入力になると考えられる。

他の IDDEX の特徴としてはドメインにおいて部品の概念が明示的に意識されていないため、計算機支援が可能となるようにドメイン知識を部品を中心に再構築したことである。上述した他のシステムのように化学構造の定式化を行ったものはあるが、部品の概念を明確にして設計知識を取り扱ったものはまだ少ないといえる。

## 7. む す び

本稿では、知的医薬品設計エキスパートシステム IDDEX について述べた。IDDEX は、スクラップ & ビルドの概念を導入することによって与えられた原材料の化合物をいくつかの部品に分解し、その部品を事例ベース推論によって置換部品を検索し、これを置換・再結合することによって得られた候補を専門家に提示するという手法を用いて医薬品設計過程の定式化と新しい計算機支援のあり方を示した。計算機による新薬設計支援を行うには、部品とその部品を取り扱う知識の充実は不可欠である。

現在筆者らは、部品獲得と並行して専門家より得られた日本語の自然言語から共通するキーワードを見だし、それを基に解析し一般化する作業を行っている。

IDDEX は、現在共同研究者であり当研究所の医薬品設計の専門家に実用されている。現在は、中枢神経系に作用する新薬を開発すべく同種の薬物数十種類を基にこれを分解して部品獲得と設計行為の説明文を入

力している段階である。専門家は、現状で設計の際の部品選択や候補提示には有用であり、今後はこれを発展させ若手設計者のための教育システムなどにも応用してみたいと評価している。しかし、IDDEX の利用にはある程度の計算機に関する知識を必要とする。今後この点についても検討していく予定である。

**謝辞** 本研究は、大阪大学産業科学研究所推進研究「知的材料設計と新素材の創出」の一部としてなされたものである。

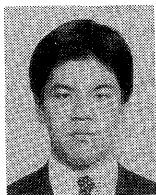
## 参 考 文 献

- 1) 岩井憲一, 金田彌吉, 植田育男, 溝口理一郎: 知的医薬品設計エキスパートシステム IDDEX について, 1991 年度人工知能学会知識ベースシステム研究会, SIG-KBS-9104-18, pp. 57-64 (1992).
- 2) 岩井憲一, 金田彌吉, 植田育男, 溝口理一郎: スクラップ&ビルドの概念に基づく知的医薬品設計エキスパートシステム IDDEX, 1991 年度人工知能学会全国大会論文集, pp. 741-744 (1991).
- 3) 関谷正明, 岩井憲一, 野村康雄, 金田彌吉, 植田育男, 溝口理一郎: 知的医薬品設計エキスパートシステム IDDEX IDDEX: Intelligent Drug-Design EXpert System, 平成 3 年度電気関係学会関西支部連合大会, G 8-28 (1991).
- 4) 三戸真俊, 岩井憲一, 野村康雄, 金田彌吉, 植田育男, 溝口理一郎: ドラッグデザインシステムについて, 平成 2 年度電気関係学会関西支部連合大会, G 8-30 (1990).
- 5) 谷田 博, 池上四郎, 奥 彬: 有機医薬品化学, 化学同人 (1989).
- 6) 山川浩司, 金岡祐一, 岩澤義郎: 新版メディカルケミストリー, 講談社サイエンティフィック (1987).
- 7) 藤田稔夫: 化学増刊 107 構造活性相関とドラッグデザイン, 化学同人 (1986).
- 8) 川田 純, 寺田 弘, 藤田稔夫: 化学増刊 116 創薬のための分子設計, 化学同人 (1989).
- 9) 岩井憲一, 金田彌吉, 植田育男, 溝口理一郎: SOAR アーキテクチャに基づく発明支援エキスパートシステム, 1990 年度人工知能学会全国大会論文集, pp. 639-642 (1990).
- 10) 岩井憲一, 金田彌吉, 植田育男, 溝口理一郎: 電子材料設計支援システムについて, 平成元年度電気関係学会関西支部連合大会, G 8-43 (1989).
- 11) 関谷正明, 岩井憲一, 野村康雄, 金田彌吉, 植田育男, 溝口理一郎: IDDEX (Intelligent Drug Design EXpert) における医薬品設計知識の獲得とその利用, 情報処理学会第 46 回 (平成 5 年度前期) 全国大会, 2D-6 (1992).
- 12) 小林重信: 設計問題へのアプローチ, ICOT/KAR-SWG (1989).

- 13) Iwai, K., Ueda, I. and Mizoguchi, R.: IDDEX: Intelligent Drug Design EXpert Based on Scrap & Build Method, *Intel. Sympo. on IDSEMS '92*, pp. 47-48 (1992).
- 14) 佐々木慎一: "Computer Chemistry" を拓く, 人工知能学会誌, Vol. 4, No. 3, pp. 240-253 (1989).
- 15) 阿久津達也, 李 春野, 鈴木英之進, 大須賀節雄: 化学エキスパート・システム構築用ツール CHEMILOG の開発, 情報処理学会論文誌, Vol. 32 No. 4, pp. 425-434 (1991).
- 16) *CSC Chem Draw User's Guide*, Cambridge Scientific Computing, Inc. (1993).

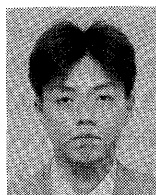
(平成 6 年 1 月 28 日受付)

(平成 6 年 3 月 17 日採録)



岩井 憲一 (学生会員)

1967 年生。1990 年関西大学工学部電気工学科卒業。1992 年同大学院工学研究科博士課程前期課程修了。現在大阪大学大学院基礎工学研究科博士後期課程在学中。医薬品設計における計算機支援の研究に興味を持つ。人工知能学会会員。



関谷 正明

1969 年生。1992 年関西大学工学部電子工学科卒業。1994 年同大学院工学研究科博士課程前期課程修了。同年シャープ株式会社入社。在学中は医薬品設計における計算機支援の研究に従事。



野村 康雄 (正会員)

1961 年大阪大学工学部通信工学科卒業。1963 年同大学院工学研究科修士課程修了。同年松下電器産業株式会社入社。同社無線研究所勤務。1970 年大阪大学研究生。1975 年同大学助手。1976 年関西大学講師。同大学助教授を経て、現在同大学教授。工学博士。電気音響変換器の CAD, 人工知能の研究に従事。電子情報通信学会, 日本音響学会, IEEE 各会員。



植田 育男

1939 年生。1962 年大阪教育大学教育学部理学科卒業。1964 年大阪大学大学院理学研究科有機化学専攻修士課程修了。同年藤沢薬品工業株式会社入社。同社中央研究所勤務。1975 年同社中央研究所主任研究員。1983 年同社部長研究員。1985 年同社副所長。1986 年大阪大学産業科学研究科有機合成薬品部門教授。現在に至る。理学博士。有機化学, メディシナルケミストリー, 創薬の研究に従事。共著書「医薬品設計—考え方と進め方—」(廣川書店)。日本化学会, 日本薬学会, 米国化学会, 英国化学会各会員。



溝口理一郎 (正会員)

1948 年生。1972 年大阪大学基礎工学部電気工学科卒業。1977 年同大学院博士課程修了。同年, 大阪電気通信大学講師, 1978 年大阪大学産業科学研究科助手, 現在同研究所教授。工学博士。パターン識別関数の学習, クラスタ解析, 音声の認識・理解, エキスパートシステム, 知的 CAI の研究に従事。1983 年度 Pattern Recognition Society 論文賞, 1987 年度電子情報通信学会論文賞受賞。電子情報通信学会, 人工知能学会, 日本音響学会, CAI 学会, 日本認知科学会, IEEE, AAAI 各会員。