

□ 最新の化合物検索様式

1. 3次元構造式検索

1.1 2次元構造式検索と3次元構造式検索との違い

3次元構造式検索は化合物の3次元構造上の特徴を検索対象とするものであり、従来行われていた化合物の2次元構造式検索(紙面上に描かれたトポロジカルな情報を基本として化合物を検索する)とは内容的におおきく異なっている。3次元構造式に限定するならば、従来の2次元構造式検索においても立体異性体(α/β 指定による立体異性対、L体及びDL体、その他)検索等が行われてきたが、その検索内容は2次元検索と3次元検索とで根本的に異なっている。3次元的情報を含む立体異性体の検索はあくまでも化合物全体が検索対象となるものであるが、3次元構造式検索は化合物内部における部分構造同志の相対的位置関係が検索対象となる。この点で立体異性体の検索は従来の2次元構造式検索技術の延長範囲で十分に実現可能であるが、3次元構造式検索には従来の検索技術ではカバーできない問題が多数存在することになる。これらの問題は、3次元構造式検索が原子や官能基の3次元的空间上の位置関係を基準として検索を行うものであることに起因する。

2次元構造式検索と3次元構造式検索のもう一つの大きな違いは、2次元構造式検索の対象は1個の化合物の集合体であるのに対し、3次元構造式検索では1個の化合物がさらに数十~数百~数千~数万という3次元異性体を持ち、これらの異性体をも検索対象にしなければならないということである。即ち、2次元検索では単体の化合物からなる1次元のデータセットだけを検索すれば良いのに対し、3次元検索では個々の化合物がさらに多数の異性体から構成される2次元のマトリクスを検索対象としなければ成らないことである。しかも、単なるリレーショナルデータベースと異なり、検索対象は全マトリクスを検索対象としなければならない。

$$2 \text{次元検索対象集合} = R^2 (X_1, X_2, X_3, \dots, X_{N-1}, X_N) \quad (1)$$

$$3 \text{次元検索対象集合} = R^3 (Y_1, Y_2, Y_3, \dots, Y_{N-1}, Y_N) \quad (2)$$

$$\begin{aligned} \text{但し、} Y_1 &= (X_{11}, X_{12}, X_{13}, \dots, X_{1, i-1}, X_{1, i}) \\ Y_2 &= (X_{21}, X_{22}, X_{23}, \dots, X_{2, i-1}, X_{2, i}) \\ &\dots \dots \dots \\ Y_N &= (X_{N1}, X_{N2}, X_{N3}, \dots, X_{N, i-1}, X_{N, i}) \end{aligned}$$

この3次元構造式検索が尋常でないことは、この発生可能な異性対の数を考えると容易に理解できる。通常、異性体の数は化合物中に自由回転可能な結合が多くなるほど指数関数的に増えてくる。例えば、自由回転する結合一個ごとに 10° 単位で回転させた時、3次元異性体(配座異性対)の数は自由回転可能な結合数を n とした時 36^n 個となる。従って、回転自由軸がわずか5個しか存在しない時でも、1化合物について約20億(2176782336)個の配座異性体が存在することになる。これだけ多数の配座異性体を計算機の中に保存しておく事は現在入手可能な最高級計算機を用いても荷が重い。なんらかの工夫をすることが必要である。ここではこのような、従来の化合物検索システムにはない特別な問題を持つ3次元化合物検索についてのべてゆく。

1.2 3次元構造式検索の実現

この3次元構造式検索技術が開発され、実用化されるには以下に示す幾つかの問題や条件が解決されることが必要である。

- (i) 化合物の3次元構造式が簡単に求められる。
- (ii) 3次元構造式検索の需要があること。
- (iii) 高速な3次元構造式検索技術が開発されること。

以下、簡単にこれらの事実について順にのべる。

- (i) 化合物の3次元構造式が簡単に求められること。

現在、化合物の3次元構造式を求める手段として様々なものがある。これらの手法の主なものとして、分子軌道法、分子力学、分子模型的基準に従った化合物の3次元組立アルゴリズムに従ったもの、3次元構造式データベース、これらの手法を組み合わせたも

の等がある。通常、これらのアプローチは解析目的により使い分けされる。

3次元構造式検索は解析を主目的とするものではない。従って、求められた3次元構造式についての高い精度は要求されず、あるていどの正確さを持っていれば良い。むしろデータベースという特性からは、多数の化合物について短時間のうちに同じ規格(精度)を持つ3次元構造式を算出出来ることの方が重要である。

前記条件を満たすものとして、従来は簡易分子力学等の計算手段を用いたアプローチがとられていた。しかし、簡易といえどもこの方法でデータベースで利用される程度の化合物数の3次元構造式を求めることは、計算時間の観点から非現実的であった。しかし、化合物の分子模型を組み立てるのと同じ感覚で3次元化合物の組み立てを行う、ルールベース主体のアプローチが取られることで計算時間は大幅に短縮された。しかも、3次元化合物検索の利用レベルに耐えるような精度で3次元構造式を求めることができる。このアプローチの出現により、はじめて3次元化合物検索の実用化が実現へと向かった。

・3次元構造式創出プログラム：CONCORD

3次元化合物データベースで最もよく利用されるプログラムがCONCORDというプログラムである。これは先に述べたように、構造式を組み立てるための様々なルールをもち、また簡易分子力学と組み合わせることで高速に3次元化合物を高速、かつ精度高く創出するものである。本プログラムによる2次元構造式を3次元構造式に組み立てる時の変換率は90%以上ある。CONCORDは有機金属化合物への適用はできない。また、ペプチド、マクロライド、ポリマーといった自由度の極めて高い化合物への適用にはむいていない。

この他にも3次元構造式データベースに利用されるプログラムとしてはWIZARD、AIMBといったものがある。これらのプログラムは総てAI的手法で化合物の3次元構造式を求めるもので、これ以外のアプローチで計算時間が高速とされる分子力学プログラムと比較すると極めて高速に3次元構造式を求めていることがわかる。

AIMB：MM2=1：110、WIZARD：MM2=1：8.9

AIMBやWIZARDといったプログラムは既に3次元構造式のわかった基本となる化合物群をデータベース中に保存し、新たに3次元構造式を求める時にはこれらデータベース中に保存されていた化合物の構造情報を最大限に生かす形で利用しつつ計算する。

従って、新たな化合物がデータベース中の構造式でうまく説明できる化合物の時、結果は良好で計算時間も少なくともすむが、データベース中の化合物と全く異なる新規の化合物が入力された時、これらのシステムから出力される構造式の信頼性は低くなる。

これらのプログラムと比べCONCORDはルールベース主体であるため、入力された構造式とデータベースとのマッチングといった問題は存在せず、平均的にすべての化合物についてかなり高い精度で、且つ高速に3次元構造式の組立を実現することができる。

(ii) 3次元構造式検索の需要があること。

技術的な環境がそろったとしても、実際の需要がなければ実用化は無意味である。3次元構造式検索の需要は、データベースを基本とした領域からの需要というよりは、実際の研究サイドからの要求からはじまったといえる。

構造-活性/物性相関の研究では化合物の3次元的な構造関係(立体化学とは異なる)を特に重視するアプローチ(例えばドラッグレセプター理論における薬物の3次元構造式とレセプターサイトの3次元形状の問題、液晶における分子の配向に関する問題等)がある。この方面での研究が進み、実際に3次元構造上での共通な特徴を持つ化合物の取り出しが現実的に必要となる場合が多くなってきている。多数の化合物群、しかも回転軸をもつことでその形が大きく変化するものから3次元的に類似な特徴を持つ化合物群を取り出すことは人間には殆ど不可能である。この仕事を計算機に行わせるのが3次元構造式検索の主目的である。

(iii) 高速な3次元構造式検索技術が開発されること。

現在、様々な形態での高速3次元構造式検索技術が実用化されている。ここでは、実際の3次元構造式検索を行うことを前提として、3次元検索の最前線を簡単にのべる。

1. 3 3次元構造式検索概要

3次元構造式検索の目的は、原子や官能基の3次元的な空間上の位置関係を満たす化合物群を探し出すことである。従って、この3次元構造式検索で必要な情報は3次元上での化合物内における原子や原子団の相対的な位置関係となる。従来の2次元構造式検索は存在情報を基本とした検索であり、3次元検索のような化合物内の位置関係の情報を扱うことはない。

実際に3次元構造式検索を行う時に必要となるのは検索キーである。この検索キーの

なかに3次元的な情報をもりこませることが必要である。従って、3次元的な情報を必要としない2次元構造式検索のキーと3次元構造式検索に用いられるキーとは大幅にその内容が異なってくるのは当然である。

1. 3. 1 検索キーについて

先にものべたように、3次元構造式検索に用いられる検索キーが備える条件は原子や原子団の空間的な位置関係を指定することが出来るという点である。この機能は従来の2次元検索には必要のない機能であり、逆に2次元検索で用いられる化合物の全体情報を指定するという機能は3次元検索では必要としない。

3次元構造式検索でヒットする化合物は、この3次元検索キーで指定された空間的な位置関係を満たす化合物である。この時、検索キーで規定されないその他の部分についての指定は何もない。この故に、3次元構造式検索の本質は2次元構造式検索における部分構造検索と同様な、検索条件を満たす化合物群を探す“類似化合物検索”が基本であり、化合物全体の一致を目指す完全一致検索ではない。

1. 3次元構造式検索キー

3次元構造式検索では原子や官能基の3次元上での相対的な位置関係の指定が必要である。このような、化合物の3次元的な位置関係を規定する一般的な項目としては以下のようなものがある。これらは、原子あるいは原子団（官能基等）を3次元空間上のポイントとして考え、原子や原子団の並び具合で規定される‘線’や‘面’といったものを基本要素として利用するアプローチがとられる。これらの基本要素を用いて指定される空間上の相対関係を以下に示す。

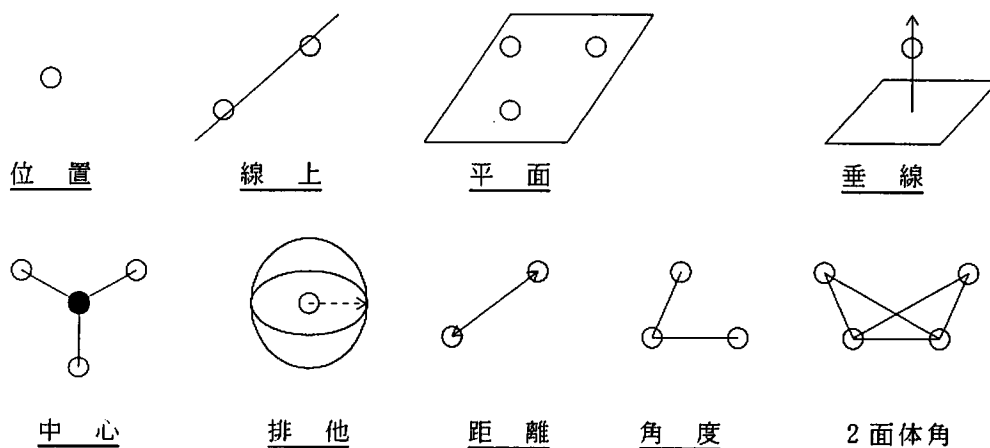
- ① ポイント：原子／原子団の位置、原子／原子団間（2原子）距離
- ② 線：複数原子／原子団の並び（線条）具合、結合角
- ③ 面：面の指定、面と面の関係（2面体角）、面と面の距離
- ④ 線と面：線と面の関係（垂線等）
- ⑤ 他：排他領域、存在領域

現在実用化されている3次元検索システムの多くはこのような基本機能を備えており、これらを組み合わせながら実際に利用する3次元検索キーを構築する。ここでは実際に実用化され、世界的に利用されているシステムを例にとり、この3次元構造式検索で利用されている検索キーについて簡単に説明する。

化合物検索システムのMACCSおよびISISを販売している米国モレキュラーデザイン社は、化合物検索システムの販売では世界の市場を独占している会社である。この会社でも3次元化合物検索システムを開発し、販売を行っている。この3次元化合物検索システムはMACCS-3DおよびISIS-3Dと呼ばれている。このシステムに関する情報は様々なジャーナル上に多数発表されているので参照していただきたい。

□ 3次元化合物検索で利用される検索キー

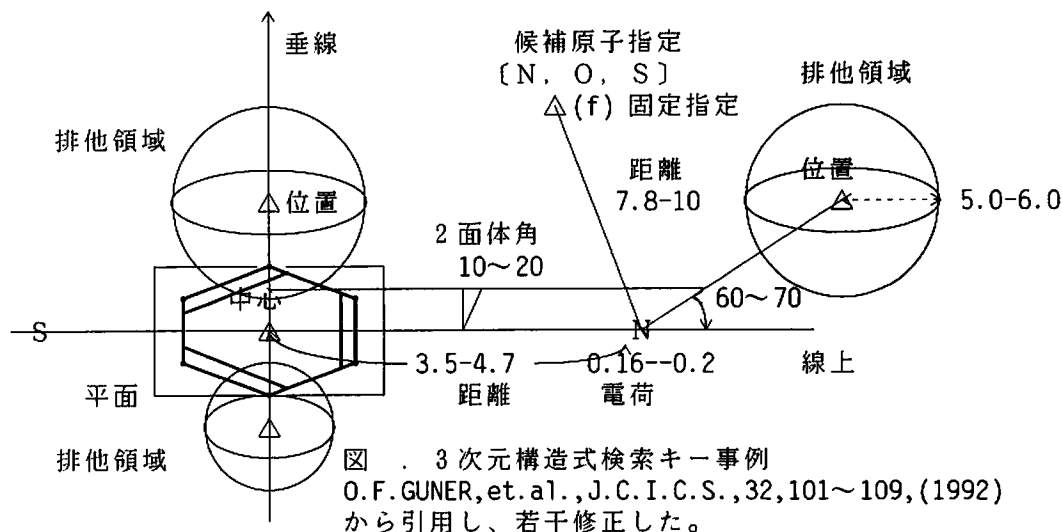
MACCS-3DおよびISIS-3Dで利用される3次元構造式検索キーを作成するための要素機能を以下に示す。



ここで示された9個の要素機能を組み合わせることで様々な3次元構造式検索のための検索キーを作成することが可能である。

これらの要素機能をより効果的に利用するために、さらに幾つかの補助機能がある。例えば、原子を空間上で固定するためのFIX指定や、個々の原子や結合の種類を指定す

るための機能、指定された各原子の適合度（指定された場所からのズレの程度）をモニタリングするためのRMS（Root Mean Square）値を求める機能等が存在する。これらの機能を上手に組み合わせて利用することで3次元検索キーが作成される。



図に前記要素機能を組み合わせた3次元検索キーの例がしめされている。この3次元検索キーには前記要素機能とその補助機能、さらには2次元検索で利用されていた機能も入っている。例えば、原子の脇にしめされた(f)のマークは該当する原子が3次元空間上で固定される補助機能を指定するものである。また[]のなかに原子が複数指定されているのは、ここの原子が[]内で指定された原子のどれかであればよいことを意味する。さらに、各原子の電荷等の情報も検索キー中に盛り込むことが可能である。また、検索キーの中で指定される数値は固定値を使うよりも、ある幅をもたせた値を用いることが多い。

この検索キーからもわかるように、3次元化合物検索キーは従来の2次元構造式検索に用いたキーと根本的に異なったものであることが具体的に理解出来る。この検索キー中、従来の2次元構造式検索で利用されてきた技術は[]でくくられた部分である。このように3次元キーは従来の2次元検索技術と組み合わせて利用されることも多い。

□ 3次元構造式検索の実行

3次元構造式検索は総てが従来の2次元検索と異なるわけではない。3次元検索の過程では従来の2次元構造式検索で用いられていた技術を利用する、ないしは2次元検索と併用するといったアプローチがとられている。MDL社のMACCS-3DやISIS-3Dは2次元と3次元検索を併用している例である。

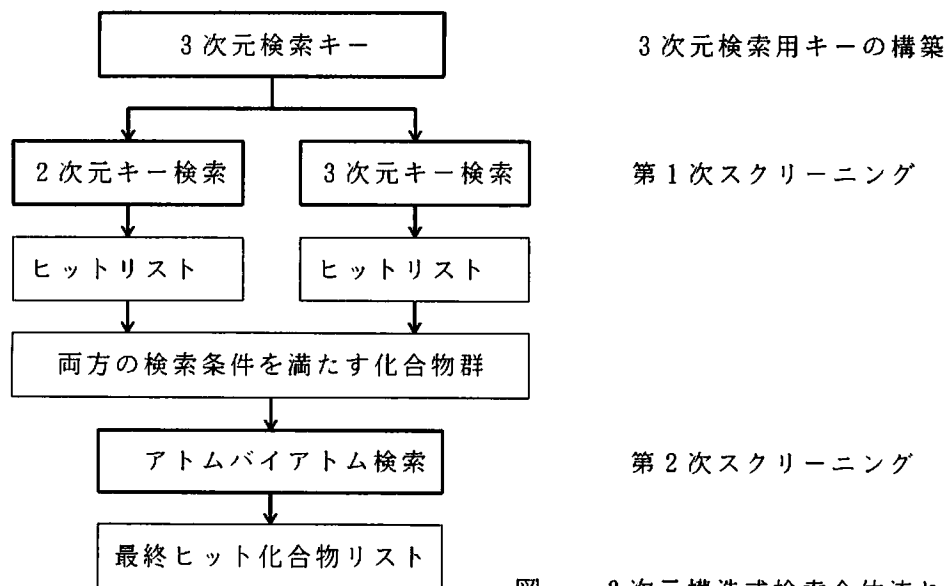


図 . 3次元構造式検索全体流れ図

3次元構造式検索の最初の段階では高速なキー検索を行い、粗いフィルタリングを行っている。この3次元キー検索に利用されるものとしてMACCS-3DおよびISIS-3Dでは3種類の検索キーを用意している。

一つは原子の情報とリングの情報を規定するための中心(セントロイド)情報に関するキーで、30種類用意されている。表にこの原子およびセントロイドに関する検索キーの一部がしめされている。

表 原子及びセントロイドに関する検索キーのリスト

1. H	6. X (halogen)	26. 7員環セントロイド
2. ローンペア	7. F	27. 芳香族5員環セントロイド
3. N	8. C1	28. 芳香族6員環セントロイド
4. O	9. O=	29. フェニル環セントロイド
5. not O	10. OH	30. 6員環ヘテロセントロイド

表には3次元検索キーで指定される原子間距離についてのキー検索用のキーを示す。

表 原子間距離の3次元キー検索用キーリスト

1. 2.2 ~ 2.5	6. 3.7 ~ 4.0	28. 12.5 ~ 13.5
2. 2.5 ~ 2.8	7. 4.0 ~ 4.3	29. 13.5 ~ 15.0
3. 2.8 ~ 3.1	8. 4.3 ~ 4.6	30. 15.0 ~ 20.0
4. 3.1 ~ 3.4	9. 4.6 ~ 4.9	31. 20.0 ~ 30.0
5. 3.4 ~ 3.7	10. 4.9 ~ 5.2	32. >30.0

この表に示されるように、各原子間の距離がその程度に応じて32の段階に分けられている。このステップは最初は0.3Å単位で区切られるが、距離が大きくなるにつれてこの区切り単位も大きくなる。30Å以上離れた原子は意味がないものとして、総てを一つのランクにまとめている。この他に結合角についても同様にランク分けされており、この結合角は $180/32=5.6$ 度単位で分けられている。

3次元キー検索はここでのべたキーを利用して、高速に検索が行われる。大体30個の原子とセントロイド、および32種の結合距離と32種の結合角の組み合わせで 2^{15} 乗種類の変化が生じる。実際のシステムではこれら総てをデータとして保存することは効率が悪いため、これらの変化のうち可能性の高い2万件をとりだし、これをさらに約2千程度のキーに圧縮して3次元キーファイルにストアしておく。3次元キー検索はこのストアされた情報をもとに実行される。

□ 3次元構造式検索の分類

現在の3次元化合物構造式検索は、検索内容から大きく2種類に分類可能である。

①リジッドサーチ

②フレキシブルサーチ

以上の2種類である。①のリジッドサーチは、検索対象の3次元構造式が固定されたものを検索するアプローチである。従って、検索によるヒットの確率は化合物がデータベース内に記録されている時の3次元形状に大きく依存してくる。すなわち、全く同じ化合物であっても、データベース内に保存されている時の構造式により、検索にヒットしたりヒットしなかったりすることが生じることになる。このゆえに、この種の化合物検索ではデータベース内に蓄える化合物の3次元構造式の吟味が必要である。

3次元構造式データベースとして大事なポイントは、同一のデータベース内では少なくとも、3次元構造式を求める基準(手法)を統一すべきであるということである。様々な基準で計算された3次元構造式を同一のデータベースに蓄えたならば、このデータベースでの検索ヒットの割合は減少することになる。また、3次元検索キーを作成する時に用いる3次元構造式創出アルゴリズムも統一すべきである。特に、自由度の高い化合物を扱う時は十分な注意が必要である。

②のフレキシブルサーチは化合物の3次元的自由度(回転可能な結合を総てフリーとし、該当する化合物が取りうる構造式のすべて)を考慮した上で検索を行うものである。従ってこの検索では①の検索で問題となった3次元構造式の異なる同一化合物の検索ミスといった不都合な問題は解決される。しかし、この検索を実行するための完全解は現在発表されていない世界上のどのシステム上でも実現されていない。現レベルではこの完全解を求めつつ、次善の策で実用化を果たしているというのが現状である。

このように、技術的には①のリジッド検索のほうがはるかに簡単であるが、一方で検索の見落とし等が発生する。また、②のフレキシブル検索は技術的に非常に高度なものを要求するが、完全に実現できるならば検索ミスの発生を防ぐことは可能となる。

ISIS-3DのCFS機能について

資料：

1. CFSについて	1
2. ISIS-3Dの検索機能	9
3. シンプレックス手法について	25
4. 文献 (206th ACS National Meeting)	
まとめ	31
本文	34

1994年 3月11日

富士通

ISIS-3DのCFISの特徴

1. フレキシブル検索に 2種類の強力なアルゴリズムを採用。
 - ・ シンプレックス法および、微分法の採用
 2. フレキシブル検索を行うのに 最適と評価されたアルゴリズム
 - ・ Willettらにより フレキシブル検索に最適とされた "Directed Tweak" アルゴリズムと 同等 (微分法) と、より 強力な (シンプレックス法) アルゴリズムの採用
 3. 多様な化合物に対して 安定性あるフレキシブル検索
 - ・ シンプレックス法で利用される関数は 多様な化合物、多様な検索オプションに耐えるものを採用。
- * 微分法は手法的制限により、シンプレックス法と比べて汎用性が少ない。

参考データ：

手法	ISIS	SYBYL	Chem DBS 3D
微分法	●	●	×
シンプレックス法	●	×	×
Systematic Search	×	●	●
U/L Search	●	●	×

* Systematic Search は微分法と比べて、検索時間は極端に大きくなる。また、ヒット化合物も多くはないという報告がある。

* Systematic Search では 環の反転ができないため、フレキシブル環がある時の信頼性は低くなる。

* U/L Searchはフレキシブル検索のプレ検索の位置づけである。

以上の事実より、ISIS-3Dのフレキシブル検索は、

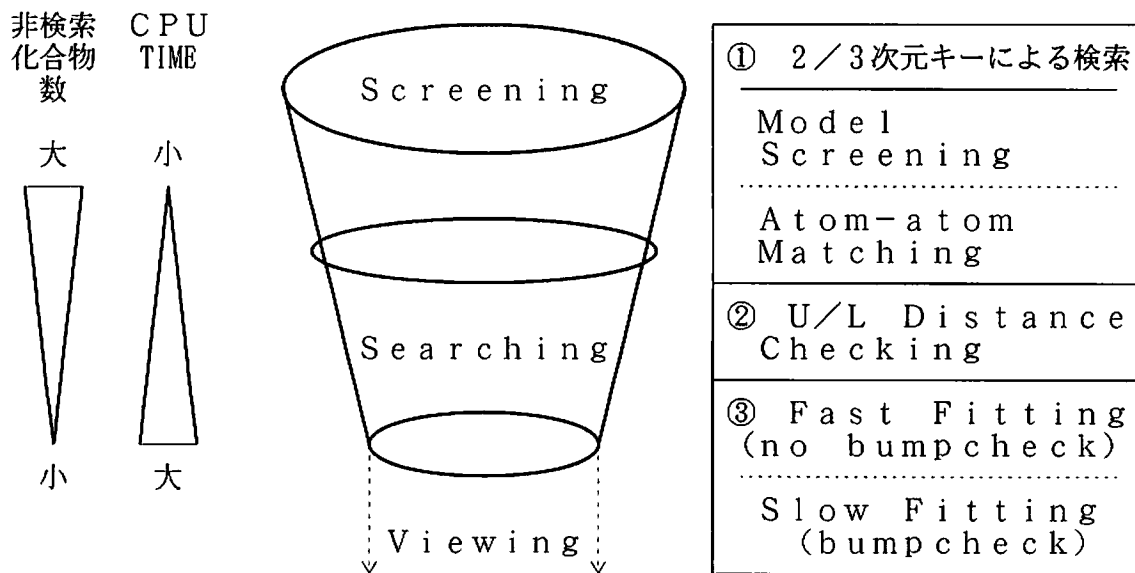
- | | | |
|--|--|--|
| | <ol style="list-style-type: none"> 1. 高速 2. 高ヒット率 3. 高信頼性 | |
|--|--|--|

といった特徴を持つと考えられる。

ISIS-3Dの配座異性体検索について

■ ISIS-3Dの3次元検索機能概要

□ 3次元検索機能構成図



* SearchingとViewingを実行する手法は基本的に同じである。違いはBumpcheckがあるか否かだけである。

□ 検索機能単位の概要説明

① 2/3次元キーによる検索 (フラグメント検索)

この検索は2/3次元構造に関する情報を種々のキー単位で保存し、あらかじめ保存されているこのキー毎のファイルを用いて高速検索を実現するものである。

* 2次元検索キー ⇨ 962キー (部分構造フラグメント)
 * 3次元検索キー ⇨ 20000キー (原子間距離と角度)

* 3次元検索キーは現在のISISでは利用していない。旧バージョンで利用していた機能である。

② Upper/Lower Distance検索 (制限検索)

2原子間の最大距離と最小距離を算出し、個々の原子が取りうる距離の最大アロウアンスを求めて、特定する。

③ Fittingを基本としたCFS

Bumpcheck (重なりチェック) の存在で、プロセスが2段階となる。

■ ISIS-3DのCFS機能

□ フィッティング機能を基本としたCFS実行

① シンプルックス最適手法によるアプローチ

② 微分最適手法によるアプローチ

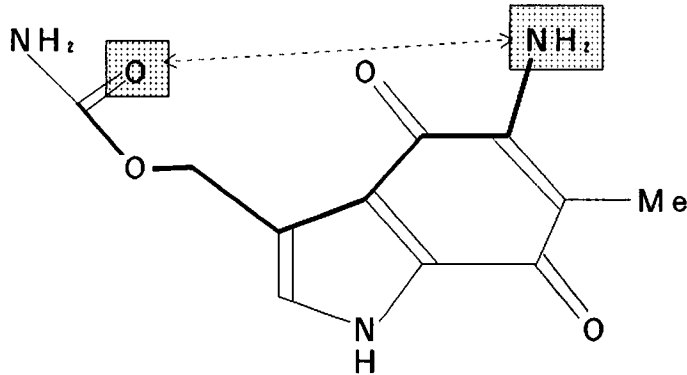
* 関連機能

- (a) U/L検索
- (b) 重なりチェック機能

■ Upper/Lower Distance Checking 事例

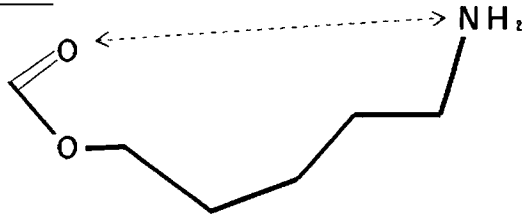
- ・原子間の最大および最小距離（最大アロウアンス）を算出する機能。
- ・プロセスは以下の5段階の過程を経る。
- ・これらの計算は検索実行時に行われる。

ステップ1



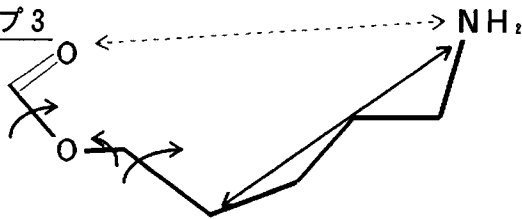
1. 原子の特定
2. 最短経路の特定

ステップ2



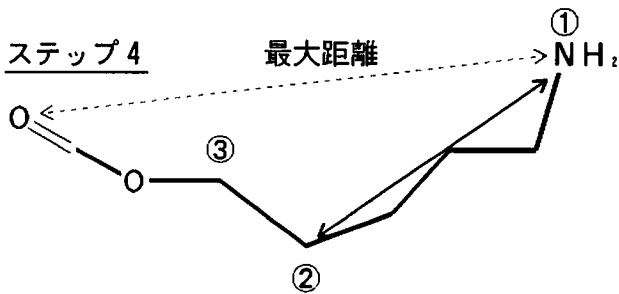
1. 最短経路の取り出し

ステップ3



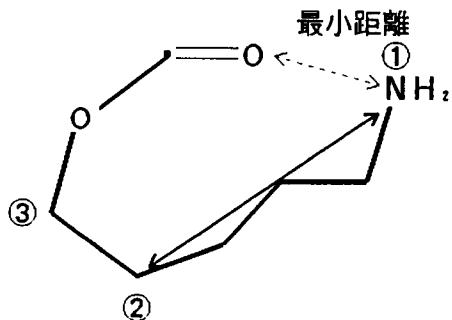
1. 回転結合および固定結合の特定
2. 固定結合をまとめて一つの結合で代表する

ステップ4



1. 回転可能結合総てについて回転させる
 2. 個々の2面体各は総て 180° にする
- * NとケトンOは①~③の平面上にくるようにする

ステップ5

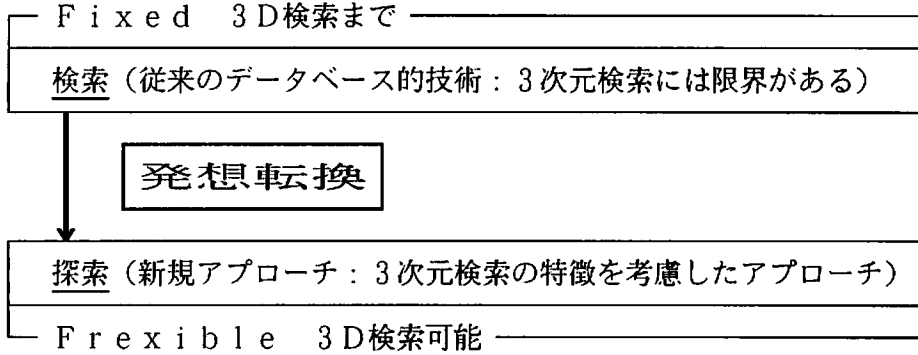


1. 回転可能結合総てについて回転させる
 2. 個々の2面体各は総て 0° にする
- * 最小距離が原子サイズ以下になった時点で終了

トーションナル検索 (CFS)

■ CFSの基本ストラテジー

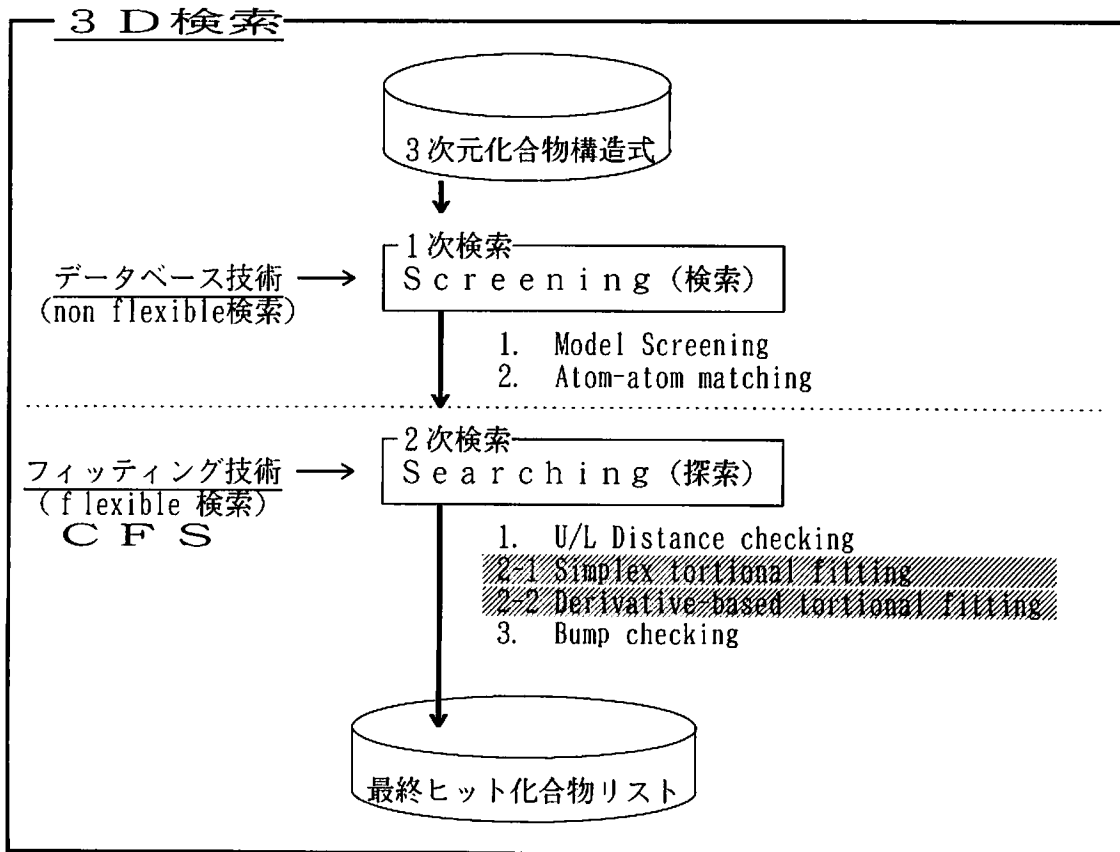
□ 検索から探索への思想変換



□ MDLのCFS検索の構成

- CFS ⇒
1. U/L Distance チェック
 2. シンプレックス法によるフィッティング
 3. 微分法によるフィッティング
- 今回追加された機能
4. Bump checking (最終チェック機能)

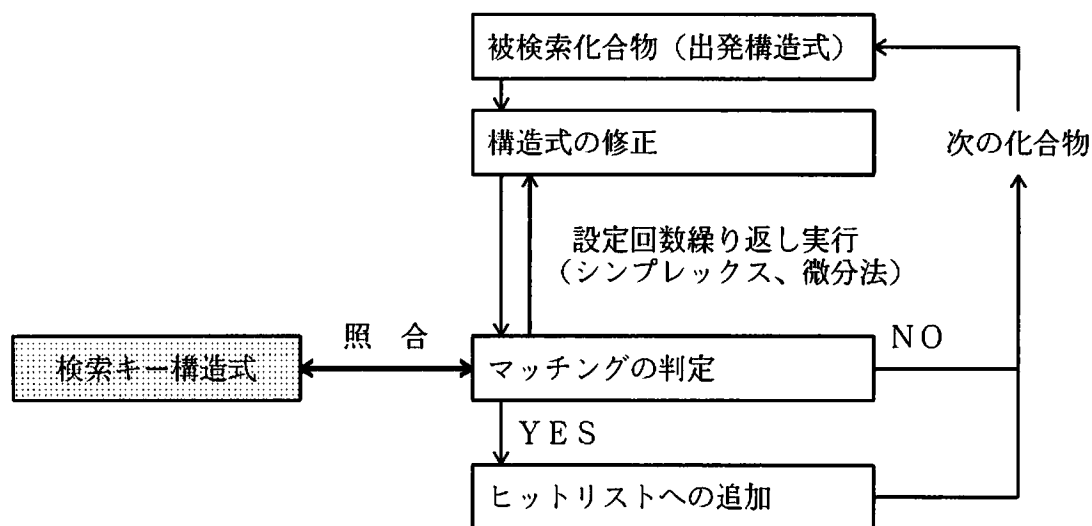
■ MDLの3D検索の流れ



■ トーショナルフィッター、またはトーショナル最適化（今回追加の機能）

被検索化合物の原子群を検索キーにフィッティングする事を基本とする。化合物のフィッティングは予め設定された関数の値の極小化過程で実行される。

□ トーショナルフィッター実行流れ図



1. 段階的シンプレックス (Sequential simplex) トーショナル最適化

本アプローチは最適化手法のシンプレックス法を用いてフィッティングするものである。この計算に用いる関数として以下のような関数を用いている。

$$\begin{aligned}
 P = & \sum_i Wd |d_i^2 - d_0^2| && \text{距離} \\
 & + \sum_i Wa3 | \cos(\theta)_i - \cos(\theta)_o | && \text{角度 (3点)} \\
 & + \sum_i Wa4 | \cos(\Omega)_i - \cos(\Omega)_o | && \text{2面角 (4点)} \\
 & + \sum_i Wl (\text{RMSline}_i) && \text{線形拘束} \\
 & + \sum_i Wp (\text{RMSplane}_i) && \text{面拘束} \\
 & + \sum_i We (N_i) && \text{排斥球 (N: 球 i 中の原子数)} \\
 & + Wf (\text{RMSfixed atoms}) && \text{固定原子} \\
 & + \text{Bump Check Penalty} && \text{非結合}
 \end{aligned}$$

* Bump Check Penalty 式

$$\text{BCP} = \sum_m Wm \frac{r_{o,i} + r_{o,j}}{r_{i,j}}$$

前記関数を用いてシンプレックス法により最適化を行う。

2. Derivative-Based 最適化 (微分法)

本手法は以下の2段階から構成される。

①最急降下法等による最適化

関数の一次微分による最適化である。本手法は極小点および最適点近傍における収束に欠点がある。この点の改良が種々試みられ、Derivative-Based最適化はその一つである。

Derivative-Based最適化は極小および最適点に近づいた時、つぎに示すBFGS法に移ることを特徴とした2段階アプローチである。

②BFGS手法による最適化

関数の1次と2次微分による最適化である。

□ 最適化に用いられる目的関数

$$P = \sum_i W_i (d - d_i)^2 \quad \text{距離}$$
$$+ \sum_j W_j (\theta - \theta_j)^2 \quad \text{角度}$$
$$+ \text{Bump Check Penalty} \quad \text{非結合}$$

*Bump Check Penalty式

$$BCP = \sum_m \left[a_{i,j} \frac{A_{i,j}}{r^{12}} + b_{i,j} \frac{B_{i,j}}{r^6} \right]$$

*メモ：シンプレックス法について

シンプレックス法は多次元データの最適化手法として著名である。微分法等に比べて種々の利点があり、広く利用されている。以下にその大きな特徴を示す。

① 目的関数に大きな制限はない。

*微分法で用いる目的関数は連続関数であることが必要

*種々の問題に対する適合性が高く、工学分野等での実績が多い。

② ローカルミニマからの脱出等の操作を行いやすい。

*最適化問題ではローカルミニマからの脱出ということも必要となる。この脱出操作や、別ルートによる探索等に関する操作性が高い。

■ シンプレックス法と微分法との比較 (T. E. Moock, et. al., J. C. I. C. S., 1994の文献より)

検索キー ID	3DSSS		CFS			
	hits	CPU	シンプレックス法		微分法	
			hits	CPU	hits	CPU
1	73	9	365	188	408	135
2	104	4	230	93	244	92
3	208	13	463	162	503	166
4	400	11	1419	699	—	—
5	10	1	116	30	120	36
6	71	2	153	15	141	17
TOTAL	866	40	1327	488	1416	446

実験条件: DEC VAX6610
Comprehensive Medicinal Chemistry database 5347化合物

考察:

1. SCREENING (3DSSS) と SEARCHING (CFS) とのCPU比較

	CPUTime	ヒット化合物数
3DSSS	40	866
CFS	467 (平均)	1372

2. シンプレックス法と微分法との比較 (ヒット化合物および検索時間)

	CPUTime	ヒット化合物数
シンプレックス法	488	1327
微分法	446	1416

3. シンプレックス法と微分法との能力比較 (被検索化合物の構造式との関係)

	評価関数	CPU*	ヒット化合物*	汎用性 (予測)
シンプレックス法	8項	大	小	大
微分法	3項	小	大	小

*1: 大きな差異はない

■ Bump check について

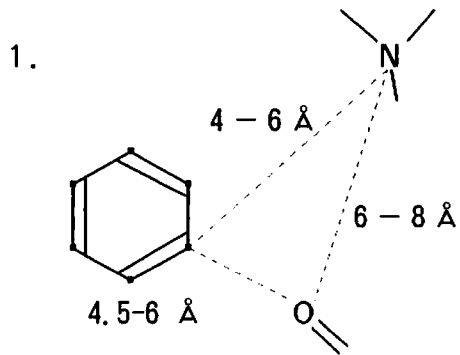
- ・ Bump check は原子の重なりをチェックする機能である。
- ・ Bump check が有る時は、無い時に比べCPUで2~4倍必要。
- ・ Bump check 検索が無い時の特徴。
 - ① 検索ヒットは広範囲になる。
 - ② 意味のないフィッティングが行われることがある。
 - ③ 検索時間はBump checkがある時と比べて、格段に早くなる。

□ 通常の検索条件下ではBump checkの有無により、2段階のCFS検索を行うことが必要である。

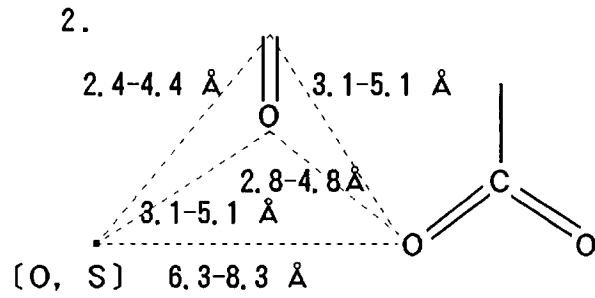
- ① CFSにおける1次検索としてBump checkのない検索
- ② 2次検索としてBump checkを行った検索

(シンプレックス法) 検索キー	Bump checking (CPU)		比 有り/無し
	有り	無し	
1	188	58	3.2
2	93	29	3.2
3	162	45	3.6
4	699	235	3.0
5	30	8	3.8
6	15	n/a	

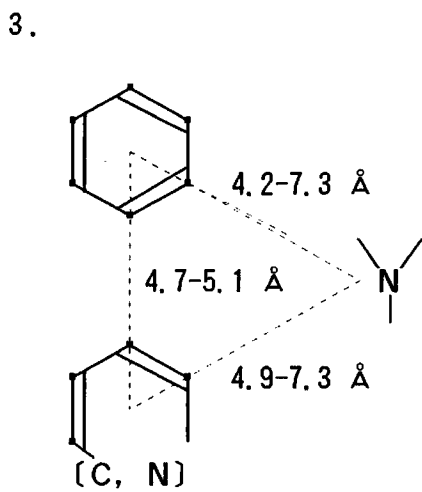
□ シンプレックス法及び微分法の比較に用いられた3次元検索キー構造式



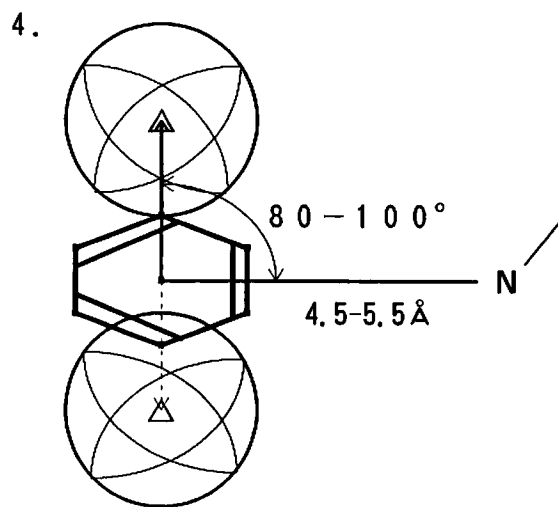
Antiarrhythmic



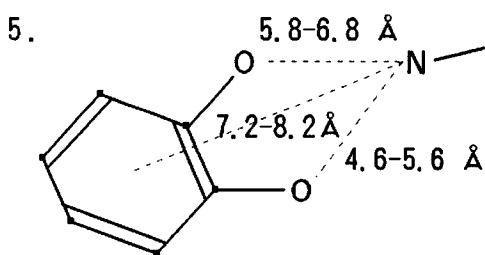
ACE Inhibitor



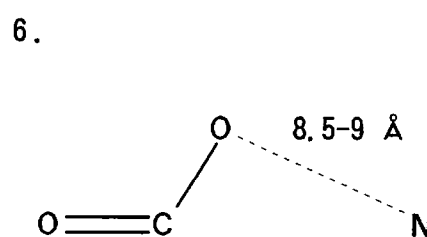
Antihistamine



Central Nervous System



Dopamine Agonist



Thromboxane Synthetase Inhibitor

Pharmacophoric Pattern Matching In Files Of Three-Dimensional Chemical Structures 'Comparison Of Conformational-Searching Algorithms For Flexible Searching'

David E. Ciark, Gareth Jones and Peter Willett, 206th ACS National Meeting

1. ペーパー概要説明

5種類のフレキシブル検索アルゴリズムについての比較報告である。以下に実験に
用いられた5種類の検索アルゴリズムを列挙する。

- | | |
|----|-----------------------|
| ① | Distance Geometry |
| ② | Systematic search |
| ③ | Random search |
| ○④ | Genetic search |
| ○⑤ | Directed-tweak search |

テスト結果では④と⑤のアプローチが実用的レベルでの検索が可能という結論が導
かれている。

尚、MDLが採用している微分法 (Derivative-based Torsional Fitter)の基本ア
ルゴリズムは⑤のDirected-tweak search に似たものである。

2. フレキシブル検索の実験条件

実験条件:

使用データベース 1) POMONA 89 データベース 1538 化合物
" 2) CAS データベース 9886 化合物
検索キー) 8種類 総て距離条件のみ

3. 手法毎の実験結果のまとめ

	CPU	ヒット率	その他の特徴
Distance Geometry	大	低	低分子不適
Systematic search	大~中	中~低	環反転不可
Random search	大	中~高	高の時は時間がかかり過ぎ
Genetic search	小	高	ポテンシャル高し
Directed-tweak search	小	高	総合点ではGAより良い

*MDL社の微分法 (Derivative Based アプローチ) は、Directed-tweak search と手
法的な基本は同じである。

以下に各手法毎の特徴を1行メッセージとして示す。○は手法の利点であり、×は欠点である。

① Distance Geometry

- ×小及び中サイズの分子のフレキシブル検索にはむいていない。
- ×時間がかかりすぎる

② Systematic Search

- かなり高い率でヒットする。
- ×時間がかかりすぎる
- ×本検索ではリングの反転が出来ない
- ×トーションの増分を小さくすると、急速にCPU時間が増大する。
 - * 被検索化合物のうち少なからぬ化合物がCPU時間制限(240秒/化合物)に引っ掛かり、検索が中断された。
 - * トーション増分が20°の時、最大のヒット率となった。

③ Random Search

- 検索ヒット率等は生成例を増やすことで、②の手法のヒット率を大きく上回る。
- ×CPU時間がかかりすぎる

④ Genetic Search

- 高いヒット率を実現
- CPU時間がかからない

⑤ Directed-tweak Search

- 高いヒット率を実現
- CPU時間がかからない
- * 全体的に④よりも高い(大きな差ではない)水準にある。
- * ④との比較は以下の理由で注意が必要である。
 - (a) ④と⑤の差異は大きくはない
 - (b) ④のプログラムは研究版で改良が進んでいないが、⑤は業務版で改良のための投資がなされた完成版である。

4. 結論

- (1) フレキシブル検索には⑤のDirected-tweak法が最適である。
- (2) ④のGenetic法と⑤の手法間に大きな差異はなかった。他の3手法は実用的観点からは失格であった。
- (3) 今回の検索は距離のみのデータを基本として作成された検索キーを用いて行われたが、実際はその他の様々な条件を加味して行われるのが普通である。この時④と⑤で次のような手法上の差異がでる。
 - (a) ④は種々の検索条件についても適用可能な手法である。
 - (b) ⑤は距離以外の条件を考慮するためにはアルゴリズムの変換が必要となる。
 - * 微分法の限界による

5. 所感

本論文では結論として⑤のDirected-tweak法を最良のフレキシブル検索法としているが、より実用レベルに近いフレキシブル検索を行う時には、ポテンシャルは④のGenetic法が高いとされている。

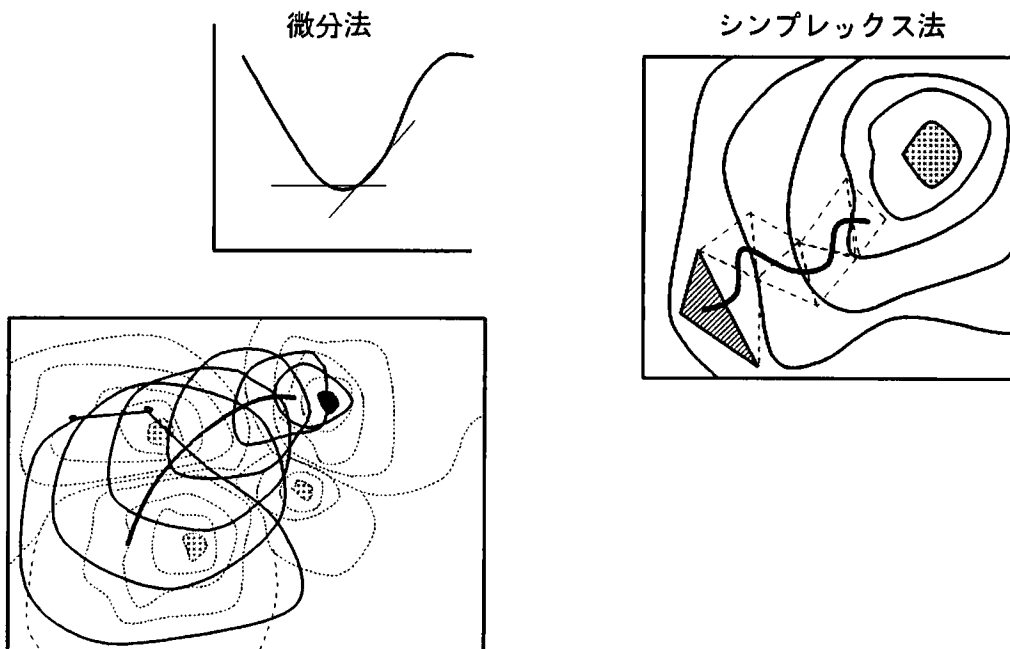
Genetic法は最近展開されてきたアルゴリズムであり、Directed-tweak法の基本となる微分法やシンプレックス法に次ぐ最新の最適化手法である。微分法やシンプレックス法が山登り法であるのに対し、Genetic法は全体をグローバルに捕らえる網掛け法である。Genetic法の優位性は種々の分野で証明されており、本論文もその一つになるものである。

□ 参考データ

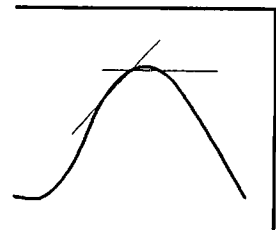
最近の最適化手法について。
最適化手法は歴史的に見るならば、微分法、シンプレックス/ダイナミックプログラミング、遺伝的アルゴリズムといった順で展開されてきた。

- ・微分法：最もオーソドックスなアプローチ
幾つかの欠点があるが、その欠点（主として極小点付近での収束の遅さ）を克服する改良法が多数提案されている。
- ・シンプレックス法：微分法につづいて提案された新アプローチ
微分法の最大の欠点である最適化関数上での制限事項をクリアしたアプローチである。種々の点で微分法を凌駕する機能を持つ。本アプローチに対しても幾つかの改良法が存在する。
- ・遺伝的アルゴリズム：遺伝のメカニズムを基本とした最新の最適化手法
提唱されたのは古いですが、最近になり再注目され始めた。遺伝のメカニズムを取り入れることで、環境に生体を自動的に適合化（最適化）するという機能を利用しようとするものである。
最近展開されたばかりで適用事例は少ないが、従来の手法にはない様々な特徴を持つことが示されている。

□ 各最適化手法の基本概念イメージ図

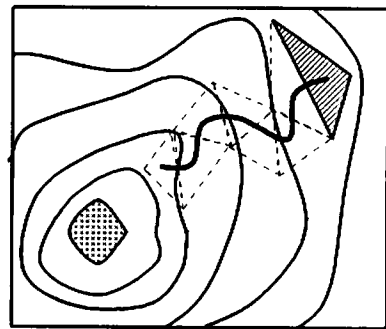


1. 微分法



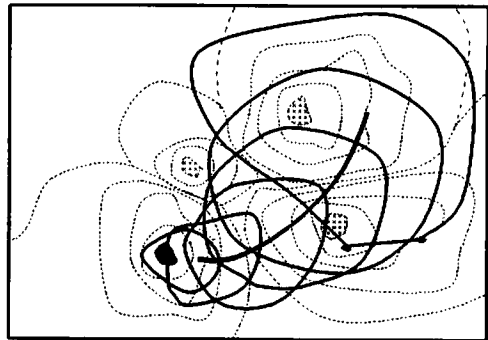
非線型関数上の接線の傾きを調べながら、極小および最小値を求めるアプローチ。

2. シンプレックス法



応答表面上のシンプレックス(三角形)をより適切と認識される方向に反転し続けることで、極小および最小点を見いだすアプローチ。

3. 遺伝的アルゴリズム



多数のサンプリング点から目的に関し優良とみなされるものを取り出し、そのサンプリング間で情報を交換することでさらに形質の優れたサンプリングを形成する。この繰り返しにより、最も優れた形質を持つサンプリングを繰り返し出すアプローチ。

□ 3次元構造式検索事例

3次元構造式検索と2次元構造式検索との基本的な差異を説明した。ここでは実際にMACCS-3D、ISIS-3Dを用いて3次元構造式検索がおこなわれた事例を以下の文献に従って紹介する。

* Osman F. Guner, Douglas R. Henry, Robert S. Pearlman, J. C. I. C. S., 32, 101 (1992)

3次元構造式検索は従来の2次元構造式検索とは大分ことなることは明白である。従って、本事例の検討は3次元構造式検索を具体的にどのように実行するかという手順(主として3次元検索キー(自在検索キー:flexible search key)の構築手法)を理解し、実際の研究にすぐ役立つようにする手助けとなろう。

○ 薬理活性および3次元構造式検索の目的

ここで行われた3次元検索の目的はACE阻害薬*1のデザインを目的として行われたものである。多くの研究者の努力により、最近になってACE阻害薬が共通に持っている必須構造要件として幾つかのものが明白になってきた。これらの必須構造要件は以下に示す3条件である。

- ① 亜鉛との結合部位
- ② 水素結合の供与基
- ③ 末端カルボン酸

これらの必須構造要件は単なる部分構造のよせ集めではなく、3次元的な相互の位置関係を主体とした情報である。従って、従来の2次元構造式検索ではこれらの要求事項を満たすことは出来なかった。これらの必須構造要件は3次元検索を行うことで初めて実現可能となり、これらの要件を満たす化合物がACE阻害薬の候補となりうる。従って、3次元構造式検索の目的はこれら3つの必要条件を満たす化合物群をデータベース中から効率良く検索してやることである。

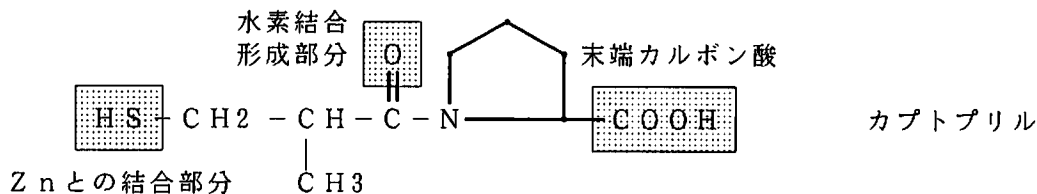
* 1 ; アンギオテンシン転換酵素 (ACE)

ACEは血中に存在するペプチドのアンギオテンシンIを血圧昇圧ペプチドのアンギオテンシンIIへと変換する酵素である。従って、このACEの活性を阻害する化合物があればその化合物は血圧降下作用を有するものと期待される。

ACE阻害薬の展開はまずACEそのものの研究から始まる。ACEの作用がカルボキシペプチターゼAと似ており、またその構造中にZn原子を含んでいることが最初に明らかになってきた。このカルボキシペプチターゼAの薬理作用の研究から、ACEの構造-活性相関も類推された。この結果、ACEには①Zn原子(Zn^{2+})と、②基質であるペプチドの末端カルボン酸と結合するための(+)電荷部分、③この2つの部分の中程にACEとその基質とをフィックスさせるための水素結合を形成するための水素原子の存在、の3項目が重要なものであることが推定されている。

ACE活性阻害のためには、これらACEの有するポケット部分に基質が入りこむことを妨げることが必要である。このため、これら3個のACEポケットと強力に結合する化合物(抗高血圧薬)の設計が積極的に行われてきた。

このACE阻害薬の最初の実用薬物としてカプトプリルが開発された。



○ 3次元構造式検索の手続き

3次元構造式検索を行うための一連の手続きを以下に順をおって説明する。

[手続き1] 目的薬理活性を有する化合物群を3次元データベース中から取り出す

既にACE阻害薬としてその活性が知られている典型的な化合物群(113化合物)を3次元データベース(16703化合物)中(この場合はMDDR*1-3D)から取り出す。

* 1 ; MDDR-3Dはモレキュラーデザイン社が同社の化合物検索システムMACCSおよびISISのデータベースとして提供しているものを3次元化したものである。データベースの内容は、最新の Patent 情報に基づいた薬理活性化合物に関するデータベースであり、Drug Data Reportのジャーナル上に報告されているものである。

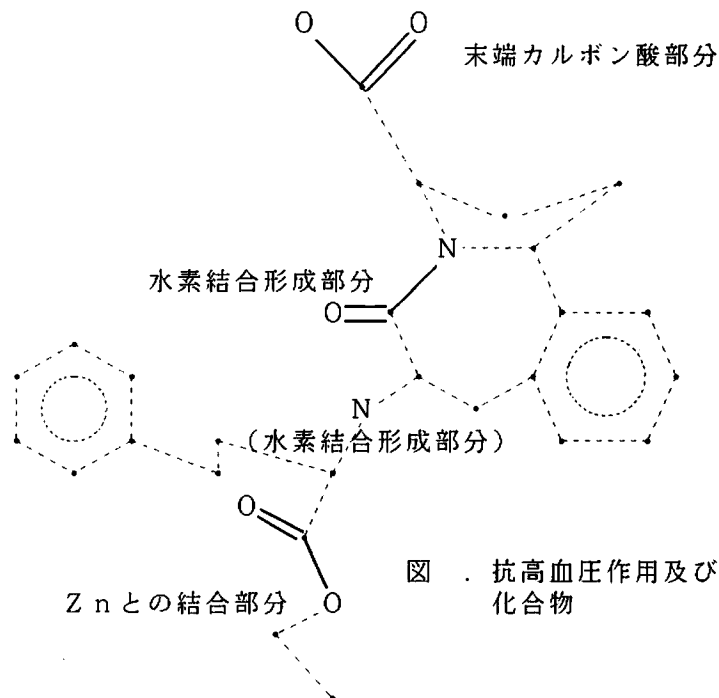
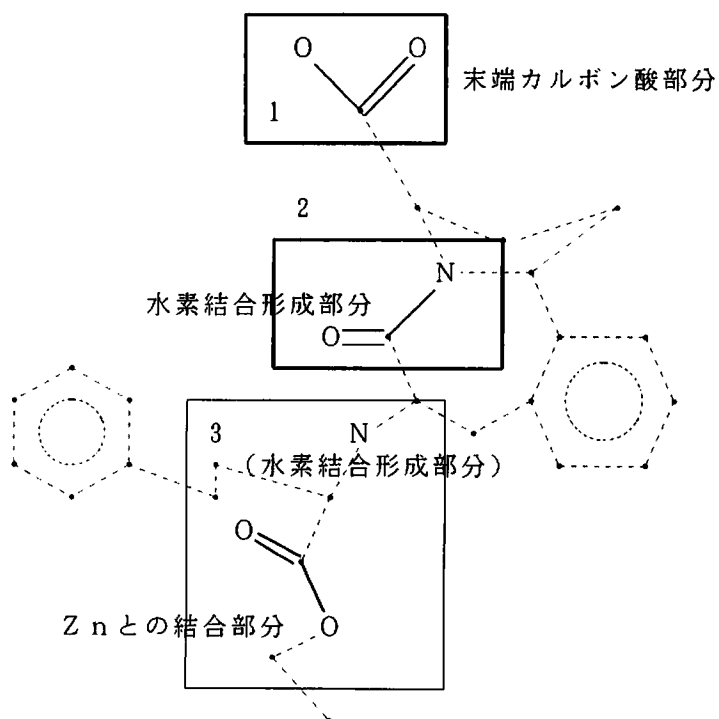


図 . 抗高血圧作用及びACE阻害作用を有する化合物

図 にデータベースから取り出されたACE阻害薬の1化合物を示す。この化合物中実線で書かれた部分は薬理作用団と推定される部分である。

〔手続き2〕候補化合物から検索に用いるパーツ（部分構造）を切り出す

取り出された化合物の薬理作用団部分（実線で表記されている部分）を取り出し、残る部分は消してしまう。この結果できた構造式の断片を検索キーの基本要素とする。



〔手続き3〕個々のパーツに対する考察

取り出された部分構造パーツのうち、太線の四角部分にある2つのグループは3次元上で比較的固定されていると考えられる。この理由は、この四角中の2パーツ

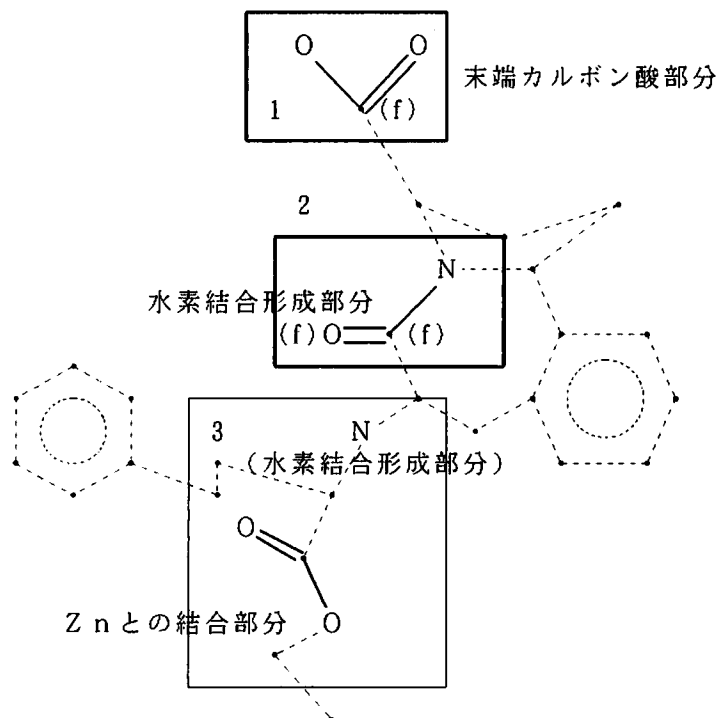
官能基)は多くのACE阻害薬中にふくまれており、しかも5及び7員環に直結、ないしは環を構成していることによる。

一方、細線で囲まれた部分は直鎖部分に位置しており、この部分の2パーツ(官能基)は3次元上での自由度は高いものと考えられる。

〔手続き4〕検索キーの細部調整

手続き3で行った考察に基づき、検索キーに対して追加条件を付加してゆく。

この事例では、末端カルボン酸の炭素、およびアミドのケトン部分を固定するための情報(f)を3次元検索キーに付加する。



末端カルボン酸の炭素原子しか固定しなかったのは、カルボン酸の炭素を中心とした自由回転による影響をさけるためである。

また、検索キーと被検索化合物との適合性をチェックする指標としてRMSの値を与えている。

表 . 全原子固定された検索及び部分固定された検索結果

TOLERANCE *1		ALL ATOMS FIXED		RIGID ATOMS FIXED	
MAX	RMS	HIT COMP.	ACE COMP. *2	HIT COMP.	ACE COMP.
0.1	0.1	1	1	6	2
0.2	0.2	1	1	3 7 6	6 1
0.3	0.3	1	1	4 3 1	6 6
0.5	0.5	3	3	6 8 0	7 2
0.7	0.7	3	3	7 9 4	7 5

*1 : TOLERANCE は許容量を示すもので、MAX は固定された(fが付いた)原子が最大取りうる距離の誤差で、RMS は総ての固定原子の移動距離に関する平均二乗根(ROOT-MEAN SQUARE)を取ったものである。

*2 : ACE COMP. はACE阻害活性化合物数を示す。

表 には検索キー中の全原子を固定した検索キーと、先に述べた原子のみを固定して作成された検索キーとを用い、RMS値を変えて検索された結果が示されている。

表から、RMS値が大きくなる(マッチングの許容度が大きくなる)につれて、ヒット化合物がふえてくるのがわかる。また、全原子を固定した時は一部のみを固

した時と比べるとヒット化合物は極端に少ないことがわかる。また、部分固定された検索ではRMSが0.2で急激にヒット化合物が増大し、その後もRMS値が大きくなるにつれてコンスタントに増大することがわかる。

一方、ACE阻害薬に関してはRMS値が0.1から0.2になる時に急激に増大し、その後は大きな変化はない。これは、ACE阻害薬の薬理作用団は3次元的に比較的固定されているためと考えられる。RMSが0.1の時はその位置的制限がきびしすぎて、様々な変動要因（例、3次元構造式創出時の誤差等）に邪魔されてヒットしなかったのであろう。しかし、RMSが0.2となることでこれらの変動要因が誤差範囲内に収まったと考える。RMSがこれいじょう大きくなると、自由度がたかくなりすぎて、目的とするACE阻害剤よりもノイズとなる化合物の方が高い頻度でヒットするようになった。

[手続き5] 検索キーのリファイン化

この検索キーをリファインすることが必要である。リファインするための手続きとして様々なものが考えられるが、最も基本的な原子間距離の指定から行う。

原子間距離の指定はある幅を持たせて行われる。基本距離±誤差範囲を与えることで検索に融通性を持たせ、ヒットの可能性を大幅に向上することが可能である。この時、原子間距離の許容は原子間の延長上にあるのではなく、固定された(fでマークされた)原子のまわりの全空間であることに注意したい。

従って、実際に検索でヒットする化合物は構成要素としては検索条件を満たすとしても、3次元的な形状が検索キーと大きく異なっているものも多くなる。

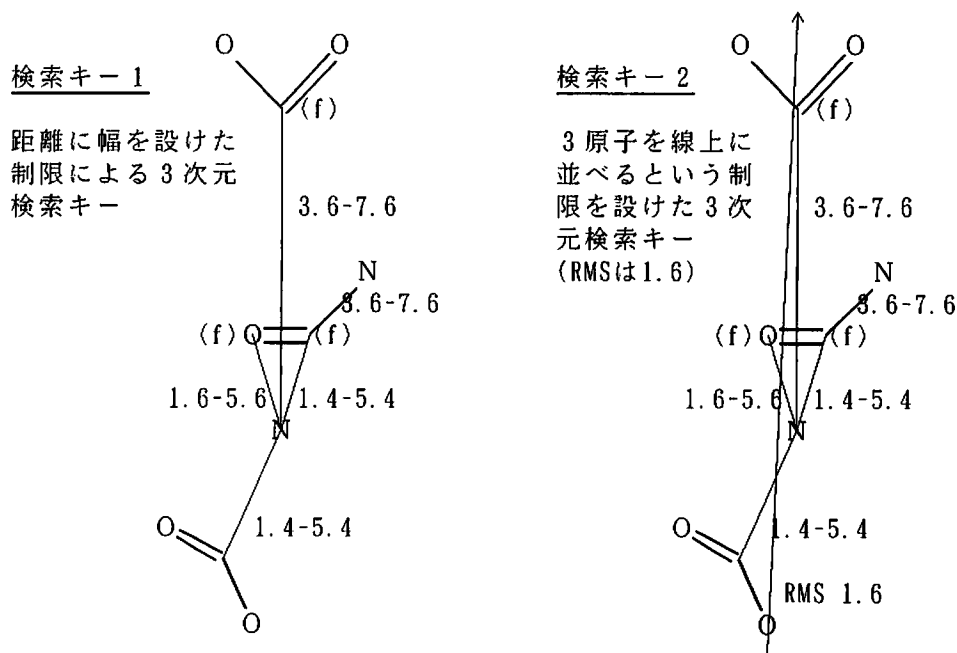


表 . 改良された検索キー 1 及び 2 による 3 次元構造式検索実行結果

QUERY ID	TOLERANCE		ALL ATOMS FIXED		RATIO(%) ACE/HIT
	MAX	RMS	HIT COMP.	ACE COMP.	
1	0.2	0.2	1 8 1	6 0	3 3
2	0.2	0.2	8 3	6 0	7 2
2	0.3	0.3	9 1	6 4	7 0
2	0.5	0.5	9 3	6 4	6 9
2	0.7	0.7	1 0 2	6 4	6 3
2	1.0	1.0	1 1 4	6 4	6 9
2	2.0	2.0	1 2 0	6 4	6 3

〔手続き6〕 検索結果を参考とした検索キーの修正

原子間距離に幅を持たせることで検索の自由度を高めることができるが、この場合3次元的自由度が高くなりすぎてノイズとなる化合物も多くヒット(181化合物)してくることになる。この中でACE阻害薬は60化合物であり、検索の精度は $60/181=33\%$ となる。

ACE阻害薬の構造式を考慮し、ここで用いられている部分構造群は3次元的に直線上に並んでいることが予想される。そこで、この制限事項を検索キーに取り入れることでノイズ化合物のヒットを減少させることが出来ると考えられる。

そこで、検索キー2のように末端カルボン酸の炭素原子とアミドの酸素原子と2級アミンを直線上でつなぐという制限をいれた。この直線上の3原子に対する許容範囲として1.6 ÅのRMS値を設定している。この新たな制限をいれた検索キー2を用いて3次元構造式検索をおこなった結果が表に示されている。

検索キー1による検索と比べると、ヒット化合物の総数が半分以下になって、しかも目的とするACE阻害薬はそのままか若干増えていることがわかる。検索精度の値は最高で72%にもなり、この種の検索では驚異的に高い値を示している。

〔手続き7〕 自由度増大によるLEAD化合物の検索

3次元構造式検索を効率よく行うという観点からは、先に展開して得られた検索キー2が理想的なものとなる。

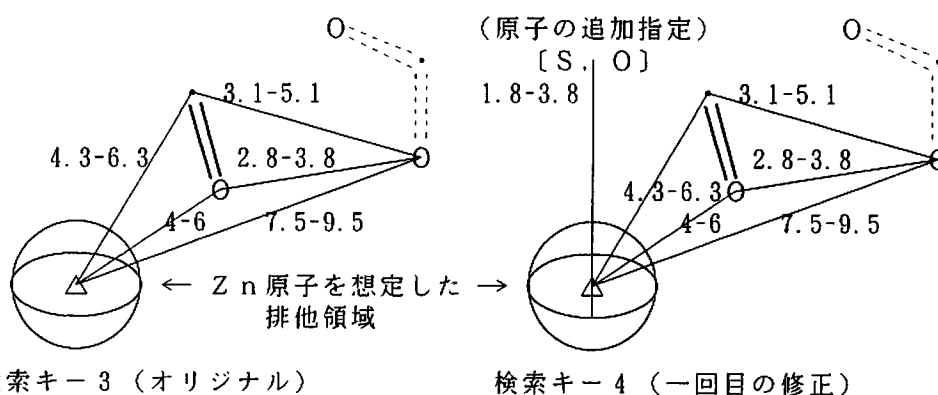
しかし、最終目的が構造-活性相関にある時はこの検索キーは最終ステップとはならない。なぜなら、構造-活性相関が目的とする新規LEAD候補化合物の選択が成されていないからである。この3次元構造式検索でヒットしたものは既に開発された既知のものであり、LEAD化合物の展開には役に立たない。

そこで、この検索キー2を用いた検索で、制限事項として付けていたTOLERANCEの値を変えることで、検索に自由度をもたせて新規LEAD候補たる化合物をデータベースから取り出せるようにする。

表でも分かるように、TOLERANCEを1.0、そして2.0とすることでヒット化合物の割合が増えていることがわかる。この間、ACE阻害薬の数は減少していない。この段階におけるTOLERANCEを増大させたことでヒットする化合物は、本来のACE阻害薬の条件を兼ね備えたものが多いはずである。勿論、TOLERANCEが小さな値にもかかわらずヒットした時のノイズ化合物もLEAD候補化合物である。

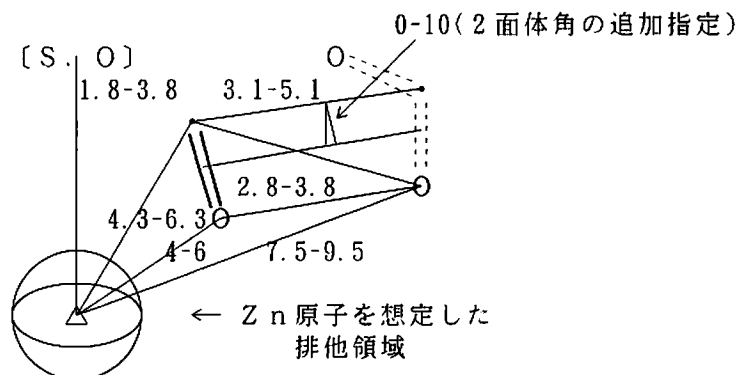
〔他の手法との比較〕

先に行った検索キーの構築は必須構造要件に基づき段階的に試行を繰り返しながら最適な検索キーを構築するものであった。このアプローチに対し、既に提唱されている薬理作用団(Pharmacophore)の構造情報を検索キーに取り込んで検索を行うというアプローチも存在する。ここではこの既に提唱された薬理作用団の情報を用いた検索を行い、前記の段階的アプローチとの結果と比較する。



ACE阻害薬の薬理作用団として、検索キー3で示されるものが提唱されている。この検索キーで排除領域となっている部分はZn原子を表している。ここでは排除領域の半径は1 Åとしている。この検索キーの中で用いられている原子間距離は既に提唱されたものを用い、その値に±1 Åの範囲を設けてある。

検索キー4は検索キー3を改良したもので、同様に検索キー5も既に発表されている情報を元に改良したものである。



検索キー 5 (2回目の修正)

表 . 段階的に構築された検索キーと薬理作用団情報を用いたキーとの検索結果

QUERY ID	TOLERANCE		ALL ATOMS FIXED		RATIO(%)
	MAX	RMS	HIT	COMP.	ACE/HIT
1	0.2	0.2	181	60	33
2	0.2	0.2	83	60	72
2	0.3	0.3	91	64	70
2	1.0	1.0	114	64	69
2	2.0	2.0	120	64	63
3	————	————	645	72	11
4	————	————	560	58	10
5	————	————	165	24	15

表からも分かるように、先に行った段階的に検索キーを構築するアプローチの方が、既に提唱されている薬理作用団情報をもちいたものよりもはるかに良い結果を出している。

検索キー 3 は ACE 阻害薬を 72 化合物取り出しているが、同時にノイズとなる化合物も段階的なアプローチによるものと比べると遙かに大きな値となっている。検索キーに改良を加えた 4 および 5 を用いた結果も同様である。検索キー 5 を用いた結果の検索精度は若干向上しているが、ACE 阻害薬の数も大幅に減少しており、この結果は全体の ACE 阻害薬のヒット率減少という犠牲に立つものである。

〔結論〕

3次元構造式検索における検索キーの重要さというものを認識すると同時に、検索キーを作成するためのアプローチ自体も大変重要であることがわかる。

単に検索結果だけから判断するならば段階的検索キーの構築が薬理作用団の直接導入による検索キー構築よりも優れていることになる。しかし、検索によって拾われてきた ACE 阻害薬の内容は段階的アプローチによる検索結果の化合物リスト中に、薬理作用団導入の手続きを経て得られた化合物リストの全化合物を含んでいるわけではない。実際、検索キー 6 のヒットリスト 58 化合物中、16 化合物 (28%) は検索キー 4 のヒットリスト 60 化合物中には存在しない化合物であった。この点で構造-活性相関を行うという立場からは留意が必要である。

ここで述べられた 2 種類のアプローチは互いに排他的に行われるものではなく、両手法を適用し、互いに相関しながら結果をリファインして行くというアプローチが取られるべき最終のものであろう。

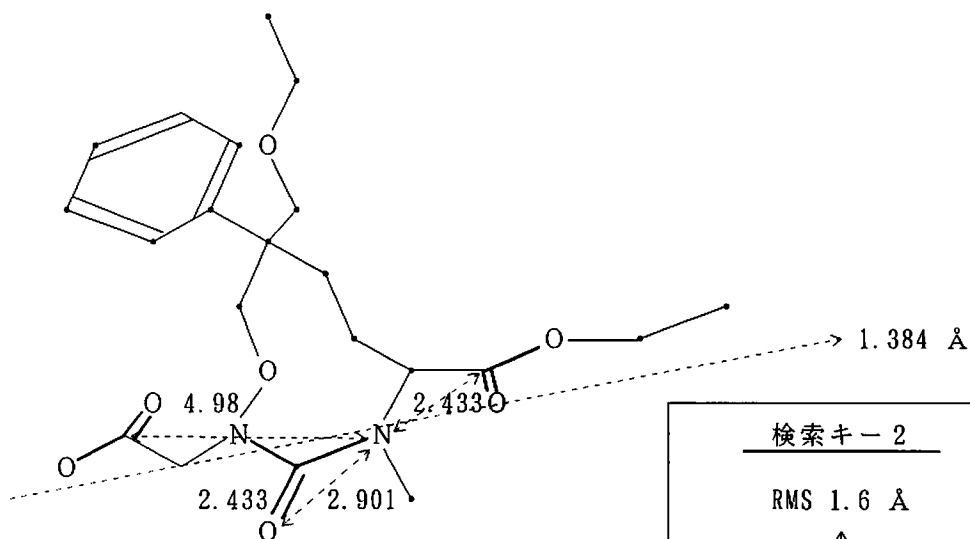
*MDDR-3D データベース (16703 化合物) 中の ACE 阻害薬として登録されている化合物数は 133 化合物である。

段階的キー構築法 60 化合物

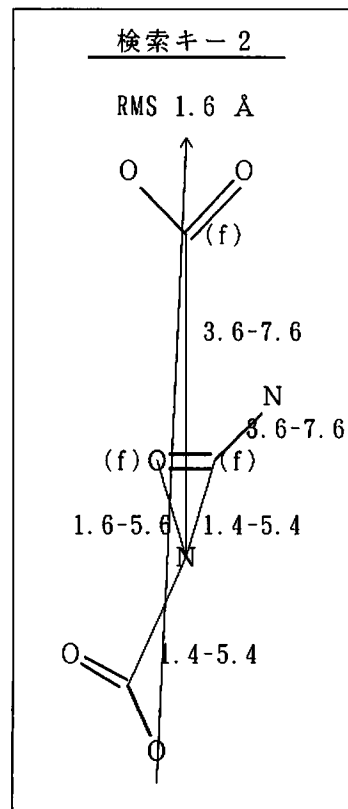
薬理作用団構築法 58 化合物

} 両手法の採用 60 + 16 = 76 化合物

[参考例 1]



検索キー 2 による検索ヒット化合物例



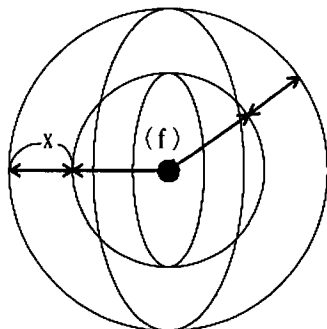
[参考例 2]

□ 固定及び範囲指定による検索範囲例

1. 一点指定の時の検索範囲

中空の球部分指定

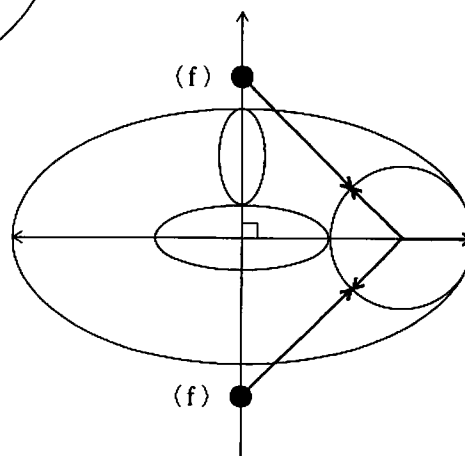
固定点 (f) を中心として全空間が指定範囲となる。



2. 2点指定の時の検索範囲

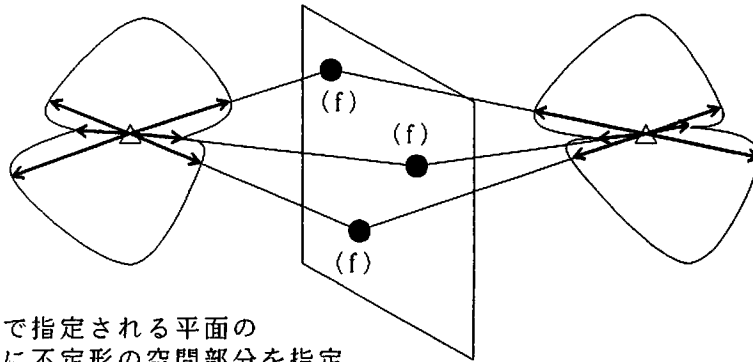
ドーナツ状部分指定

固定された2点を貫く垂線のおおる面上におかれたドーナツ部分の全空間が指定範囲となる。



3. 3点指定の時の検索範囲

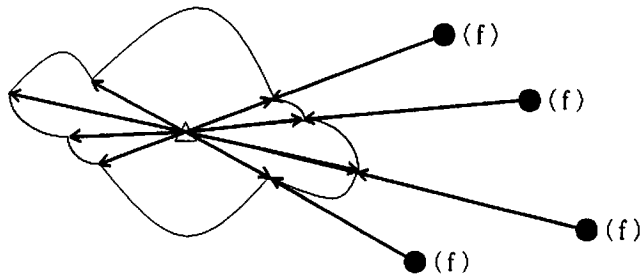
不定形部分指定



3固定点で指定される平面の
両サイドに不定形の空間部分を指定。

4. 4点指定の時の検索範囲

不定形部分指定

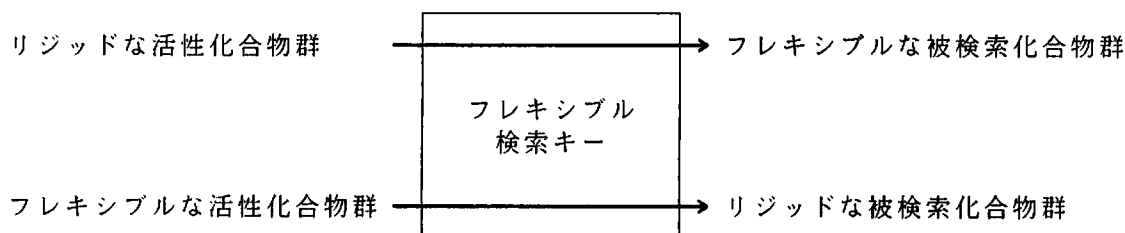


固定された4点の指定範囲の大きさで不定形部分が1ヶ所ですむのか2ヶ所となるのかがきまる。3固定点できまる平面の両側の不定形部分が4点目の範囲に入るか否かで1ヶ所および2ヶ所が決定される。2ヶ所指定された時、その形は面をはさんで非対象となることが多い。

図には1ヶ所指定部分の例を示してある。

□ フレキシブル検索キーによるリジッドな骨格構造の探索

フレキシブル検索キーはその検索対象をより自由度の高い化合物とするが、検索キー作成のための化合物を自由度の高い化合物とし、被検索化合物群をリジッドな化合物群とすることも可能である。このように、従来のパターンと異なる形態の検索要求は、自由度の高いペプチド系統やマクロライド等の活性化化合物群を出発化合物とした構造-活性相関等で発生する。



□ 検索事例

74個のアミノ酸からなるペプチド (t e n d a m i s t a t) を出発化合物として、3次元検索キーを作成する。この検索キーを用いて3次元個検索をおこない、ヒットする化合物中リジッドな基本構造式を持つ化合物を新たなLEADターゲットとする。

このペプチドは α -アミラーゼの阻害作用を持つ。このペプチドを基本とし、最終的な標的化合物としては同様な活性を持つリジッドな化合物を探すことである。

このペプチドの活性部位はTrp₁₈-Arg₁₉-Tyr₂₀であることがわかっている。この活性部位を構成する3アミノ酸のうち、個々のC α -C β 結合を検索キーとして3次元化合物データベース(ケンブリッジX線結晶解析データベース)の検索に用いた。¹⁾

1) Bartlett, P. A.; Shea, G. T.; Telfer, S. J.; Waterman, S. CAVEAT: A Program to Facilitate the Structure derived Design of Biologically Active Molecules, in Molecular Recognition: Chemical and Biological Problems, Roberts, S. M.; Ley, S. V.; Campbell, M. M. Eds. 1989, Cambridge, Royal Society in Chemistry, pp 182-196.

ここで行う3次元構造式検索では、検索キーを作成するにあたっての基本的アイデアはBartlettらが採用したものをそのまま利用し、実際に検索キーを作成したり、検索を行う時のデータベースとしてはMDLが提供するFCD-3Dを用いて実験を行っている。

採用された基本アイデアとは、活性部位とみなされるTrp₁₈-Arg₁₉-Tyr₂₀の3アミノ酸におけるC α 及びC β 位の炭素原子を指定し、これらを用いて検索することである。この大きな理由は、C α 及びC β 位の炭素原子はアミノ酸の他の部位の原子(極端に言えば末端原子等)と比べて自由度が低いと考えられるからである。

[手順1]

最初に3次元検索キーを作成するための基本化合物を作成することが必要である。CONCORDを用いて、先のTrp-Arg-Tyrの3アミノ酸の3次元構造式を創出した。ここで創出された構造式を用いて3次元検索キーを構築する。

*ここで3次元検索キーを構築するのに用いた化合物の構造式はCONCORDで創出された構造式を用いている。このことは、3次元検索でヒットする化合物は必ずしも真のLEAD候補化合物とは言えない。

正確にLEAD候補化合物を検索するならば、3次元検索キーを作成する時に利用される化合物の構造式は活性部位に結合した時の活性構造を取っていることが必要である。この検索作成用の化合物が活性構造に近ければ近い程、3次元構造式検索によりヒットしてくる化合物の構造式は真のLEAD化合物に近づいてゆく。

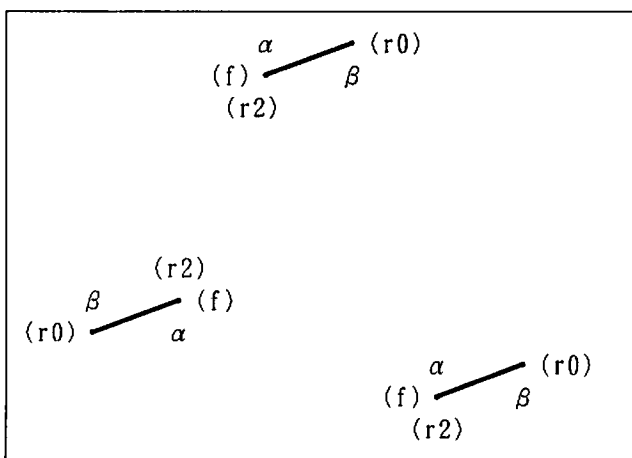
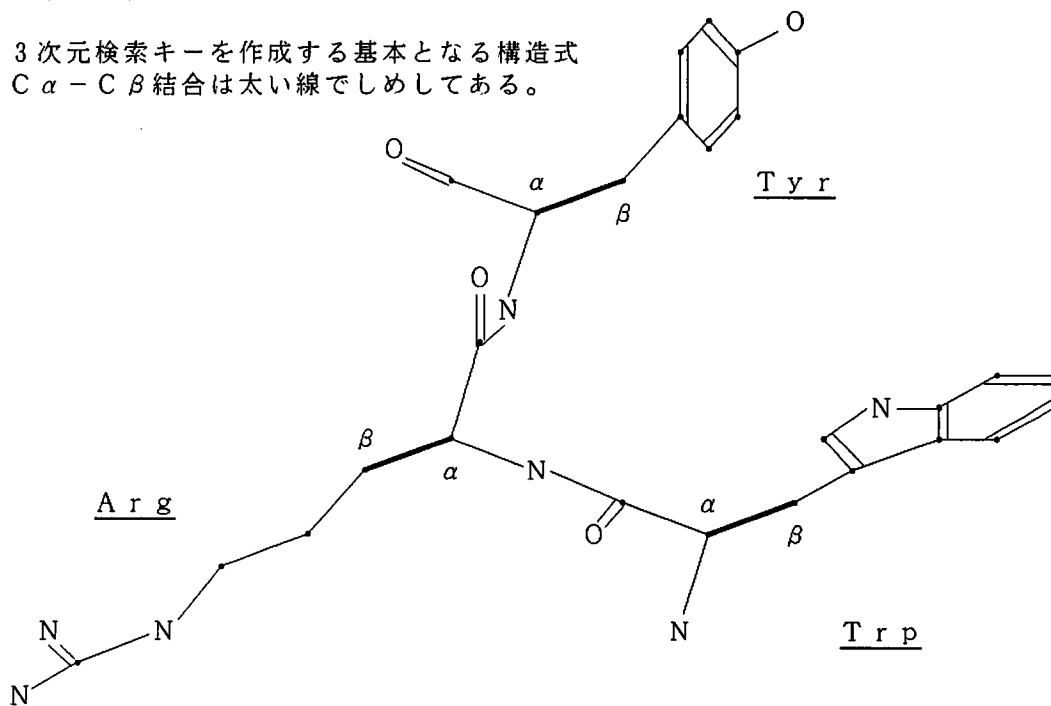
[手順2]

創出された構造式中 α 炭素と β 炭素とを取り出し、残る部分を消去する。

この取り出された炭素原子に対し、 α 炭素はリングの一部を形成する(リングの原子に最低2個つながっている)という制限を設け、 β 炭素はリングシステムに組み込まれ内という制限を設ける。

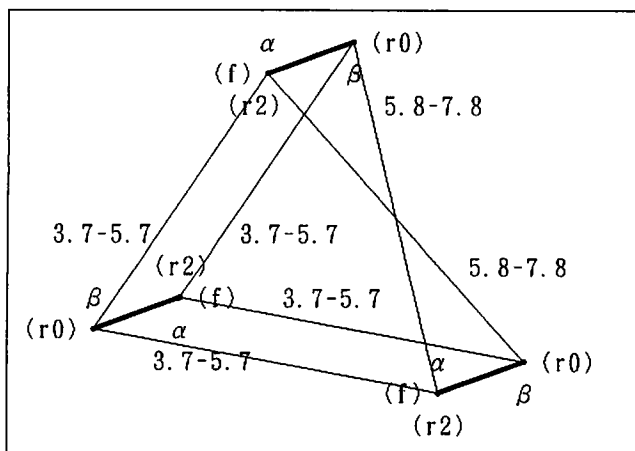
[基本化合物構造式]

3次元検索キーを作成する基本となる構造式
C α -C β 結合は太い線でしめしてある。



[3次元検索キー 1]

- * (r2) : リング原子に最低2個結合しているという指定
- * (r0) : リング原子に結合することはないという指定
- * (f) : 原子固定指定
- * α 、 β : α 炭素および β 炭素

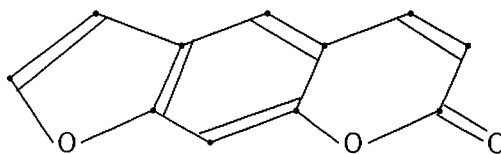


[3次元検索キー 2]

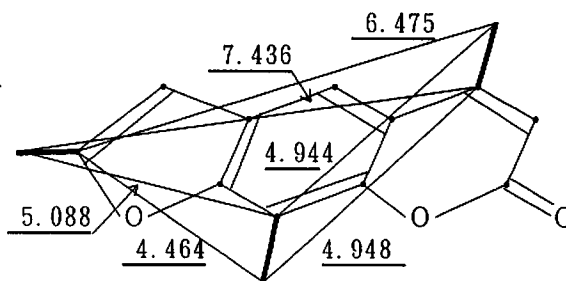
自由度の高いベータ炭素に制限を設ける。
 β 炭素は自分以外の2個の α 炭素と互いに結合される。この時の結合の自由度の幅は ± 1 Åとしてある。また、固定された3原子のRMS値は0.5 Åとした。
この検索キーを用いてFCD-3Dのデータベースを検索。検索結果5化合物がヒットした。

[検索ヒット化合物]

最終ヒット化合物構造式



最終ヒット化合物と
3次元検索キーとを
重ね合わせ表示



1. 3次元データベース

実測データによるデータベース

① C C D C (Cambridge Crystallographic Data Centre) の C S D (Cambridge Structure Database)

C S Dは1993年現在約10万件の有機化合物(低分子)のX線解析による結晶データを集めている。創出された3次元データでなく、実測の3次元データであるという点で貴重なデータベースである。

3次元検索のためにはQ u e s t - 3 Dというソフトを利用する。このソフトウェアはG e o m, G e o s t a t, G S t a tの3モジュールから構成されている。

C C D Cでは今後M D LのM A C C S - 3 D用およびS y b y l / 3 D B U n i t y用のフォーマットでの提供が予定されている。

② Brookhaven National Laboratoryの P D B (Protein Data Bank)

P D Bはタンパク質や核酸の3次元データベースである。

計算による3次元データベース

① C A S (Chemical Abstracts Service)による3次元データベース

C A Sは現在約一千一百万化合物を登録している。登録された総ての化合物について3次元座標データを求めているのではなく、3次元データベースへと変換されている化合物はこの一部分である。この3次元データベースは、C A S T - 3 Dと呼ぶ。

このデータベースには約50万化合物(1993年)登録されている。そのうち約37万化合物は構造的にリジッドな化合物である。これらのリジッドな化合物中、E X O C Y C L I C な部分を持つ環状化合物が72%、その他の環状化合物が26%、非環状化合物が2%を占めている。

データ内容:

- ① C A S レジストリー番号
- ② 化合物結合表
- ③ 化合物3次元座標
- ④ 他

このC A S T - 3 DはM A C C S - 3 D, S Y B Y L / 3 D B 及びC h e m D B S - 3 Dでの検索が可能である。

3次元座標はC O N C O R Dで生成されている。

このデータベースはC A S のRegistryファイルやSTN Express やKermitを用いて直接ダウンロード可能である。

② Chapman & Hall Dictionary of

出版会社が提供する3次元化合物データベースである。このデータベースは登録されている化合物の内容に従って3種類に分けられている。これら3種類のデータベースを以下に示す。

D r u g s : 約12000化合物(薬物に限定)のデータベースである。これらの登録化合物には、薬理データ、危険及び毒性データが付随している。

N a t u r a l P r o d u c t s : 約54000化合物(アルカロイドおよび抗生物質)のデータベースである。

Fine Chemicals: 約12万化合物のデータベースである。
62000の6員環芳香族化合物、8000の5員環芳香族化合物、12000の6員環ヘテロ環化合物、5000の5員環ヘテロ環化合物、11000の縮合6員環化合物、3000の縮合5員/6員環化合物である。
これらの化合物はChemical Design社のChemDBS-3Dで利用可能である。

③ MDL社提供の3次元データベース

MDL社が提供している化合物データベースの3次元版である。

MDL社の3次元構造式は総てCONCORDを用いて計算される。

・Fine Chemicals Directory-3d: 約63000(3D)、75000(2D)化合物収録されている。世界の77の化学薬品メーカーからの提供品がリストしてある。

・MACCS-II Drug Data Report-3d: 約27000(3D)、34000(2D)化合物のデータベースである。これには世界の11の特許事務所から報告される化合物のリストである。

・Comprehensive Medicinal Chemistry-3D: 約5400の3次元および6200の2次元化合物データが入っている。医薬品として利用される化合物のリストである。