

分子図の手書き入力システム

田中伸英*

1. はじめに

我々は色々な分子計算のために必要な初期座標を入力するシステムとしてイメージスキャナによる手書入力の可能性を研究してきた。近年、パソコン能力の強化にともない大規模な分子計算プログラムも一般の研究者に手がとどくようになってきている。これらの計算には分子を構成する各原子の3次元座標と原子間の結合関係の情報が必要である(図-1)。結合関係のデータは図から簡単に読み取れる。しかし座標はそうではない。この問題に対して、グラフィックと座標データベースを利用した対話型座標作成システムの開発が幾つか発表されている。データベースに求める座標がない場合には画面上で別の分子の分割および結合などで望みの座標を組立てて行く方式である。この場合、座標入力はいくまでも余分な仕事でありこのことに多くの労力と時間を裂くことは感覚的に好ましくない。我々は違和感が少なく、出来るだけ少ない操作で、計算機をあまり使ったことがない研究者が座標入力出来る方式として、手書き図形入力方式が有力であると考えている。そして立体的な多環状の分子の手書図からその初期座標を導出するシステムの開発を行っているので現状を報告する。この報告は計算機センター特別研究プロジェクトの規定による中間報告である。

2. 方法

現在このシステムで取り扱う分子図としては環状炭化水素の線画を仮定している。さらに炭化水素はベンゼン環や多重結合を持たないものとする(図-1)。線画の角点は炭素原子の存在する位置である。水素原子は省略する。水素原子の座標は炭素骨格が決れば幾何学的方法で計算可能である。立体感を出すためと交わっていないことを明示するために骨格の後ろ側の結合は切断して書く。処理の流れは以下の三つの部分からなっている。そして(4)に示されるプログラムやシステムへの接続を考えている。以下にシステムの概略を述べる。

- (1) 原画像の入力
- (2) トポロジーと2次元座標の抽出
- (3) 簡易エネルギー計算による初期座標の導出
- { (4) 表示および各種の分子計算：他で開発されたプログラムを含む。 }

* 学習院大学理学部講師, 計算機センター所員

(1) 原画像の入力

線画の入力はパソコン(PC9801UV)に接続したイメージスキャナ(PC-IN502)を使用している。(図-1)

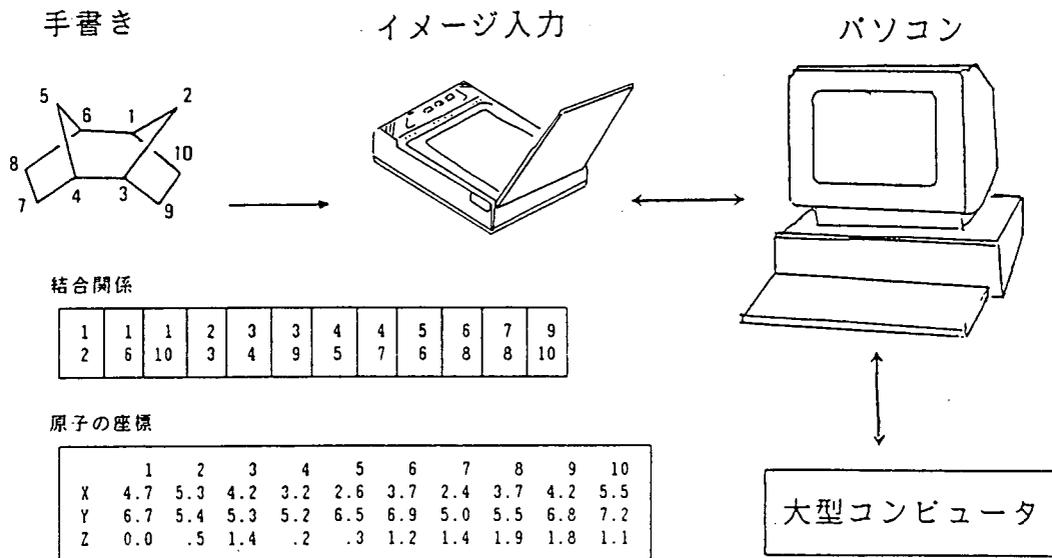


図-1 処理の流れと得られるデータ

手書き図の番号は説明のため加えてある。結合関係は縦の組が辺を表す。

プログラムは付属マニュアルに従ってC言語で作成した。画像の読取り線密度は90本/インチ(約4本/ミリ)で入力したが通常の大サイズの分子図には十分であった。入力画像は白黒の2値画像として読み取り、パソコンのVRAMへ書き出される。

イメージリダは29cm * 20cmの大きさまで読めるが次のステップで使用するプログラムのメモリの制限から現状では名刺大(9cm * 5cm)に限定される。

(2) トポロジーと2次元座標の抽出

線画から結合関係と角点の2次元座標を求める方法は以下の順である。

- ① 原画像の細線化
- ② 次数2以外の点の抽出
- ③ 次数2の点の抽出
- ④ 修正

① 原画像の細線化

VRAM上のデータを二次元配列に読み込む。データは線の存在する部分が1, 存在しない部分が0として記憶される。このままでは線分は幅を持っているので分岐や角点を取り出す処理が複雑になる。よって8点連結による細線化をおこない各線分を折れ線に集約した。細線化によって線のトレースが一意になり頂点や結合関係の発見が簡単になる。しかし細線化により角点にひげが現れたり、一つの交点分離して二つに成ることがあるがこれらの修正は最後の部分④で行う。細線化やその他の前処理には画像処理プログラムパッケージ(SPIDER)を使用して作成した。

② 次数2以外の点の抽出

この目的のためにはスタックを利用する一般的手法を使用した。この段階では細線化で現れたヒゲや分離交点そのまま結合表に書込む。このトレースに合わせて各辺ごとに辺を構成する点の座標とチェーンコードを取り出しておく。これにより以後、トレースは図を使用しないで行うことが出来るので計算スピードが上がりメモリの使用も大幅に減らせる。初めに書かれた分子図が非連結な部分に分かれる場合の処置は現在していないのでその様な図では不都合が起こる。つまりどちらかの部分がトレースされない。

③ 次数2の点の抽出

②で得られたグラフの各辺をトレースして次数2の点を探すには二つの方法を検討中である。一つはデジタル直線の直接のトレースによる方法であり、他は直線の傾きを計算して探す方法である。

④ 得られたデータの修正

得られたデータは次の二点を修正する必要がある。一つは、細線化で現れたヒゲや分離交点これらは本来存在しないはずの構造である。幸いにもこれらの線分の間隔は普通の線分比べて短いのが特長的である。よって辺の長さが一定値以下のものをデータから除いている。

二つ目は切断辺の復元である。分子図は立体感を出すためと交わっていないことを明示するために骨格の後ろ側の結合は切断して描いてある。この切断辺を復元して一本の線にするためにある長さで2端点が接近していて、間を辺が交差している構造を探して切断辺の復元をしている。

(参-1)以上2点に対するパラメーターの値は実験中である。

(3) 初期座標の導出

基本方針は(2)で取出された結合表に従い結合長と結合角にバネモデルを仮定する。そして、2次元座標から初期値を推定してバネエネルギーが最小になる原子の位置を計算する。得られた座標が他の計算の初期座標になる。極値探索にはHOOK-JEEVES法を使用した。(参-1, 2)

以上のプログラムは(1)を除いてPC-98上のMS-FORTRANで書いてある。グラフィック表示のためにユーティリティソフトHG Xを使用している。

3. 結 論

イメージスキャナを利用して対話型でない分子座標入力システムを作成した。この方式は対話型と違って多くの命令を必要としないので簡単に座標を入力できる。またスキャナの改良によっては、コピーのように入力図を何枚もセットして自動的に次々と座標の入力が出来る可能性を持っている。絵に描いた通り入力可能なので立体異性体を特徴付ける座標もそのまま計算されるのも利点である。問題点としては、角点が認識され易いような形でかつ直線をはっきり書く必要があること、極値探索の収束が失敗した場合の処理等があげられる。現在これらの点に関してプログラムの改良を進めている。

参 考 文 献

- (1) 田中, 菅, 飯塚 「グラフ構造の入力について: 分子図と3次元構造」情報化学討論会 44-46, (1985)
- (2) 田中, 今井, 菅, 飯塚 「分子構造の情報処理: 2次元分子図から3次元座標の導出」情報処理学会第29回後期大会 2M-8 (1984)