

有機化合物の構造解析における情報処理法

1Z-1

工藤 喜弘

山形大学工学部

〔序論〕 有機化合物の構造解析には化学の総合力を要するが、節目ごとの重要な判断以外は、機械化できる。ただしそれだけでは化学者の思考の循環系からぬけるため、化学者に不安感を与えるので、本研究では化学者が安心して使えるシステム作りのための要件を検討した。

〔構造解析の本質〕 構造決定の論文の中にはいったん発表した構造の修正に関するものもある。まちがいの原因がデータの解釈まちがいによることはほとんどなく、大部分は他の可能性の見落としである。それに気づくのは、たいてい、さらなる情報を獲得して矛盾が出たときである。

まちがいを防ぐには可能性全部を列挙すればよい。一般の問題では限界がまったく不明で、解の存否すらわからないのも普通である。それにたいして構造解析は(情報皆無で可能性無限のとき以外)列挙が可能な数少ない例の一つである。ただしいくら有限でも情報が少ないときには事実上無限であるから早い時点から全数の直接列挙を意図するのは現実的でないし、する必要もない。

全数列挙が可能でしかもその中に最終的な正解1個がかならず含まれていることが期待できる。そのような問題に対しては情報同族体(図1)検出法の中では「あみ法」が適している。

O 情報同族体
X 情報矛盾体

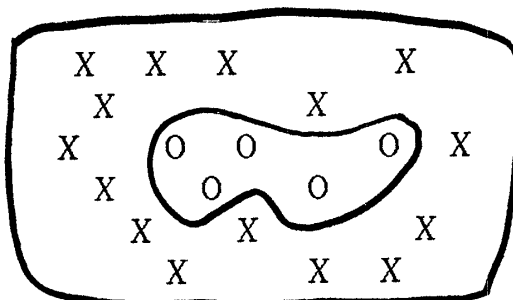


図1 情報同族体と情報矛盾体。
(間に未識別体もある)

〔情報同族体〕 情報理論はものごとの識別には一定量以上の情報が必要であることを教える。すなわち、情報が不足の場合は、手順に関係なくその情報だけでは同族とみなさざるをものの集合が発生する。それが情報同族体である。情報同族体検出のための代表的な手順が「つり法」、「あみ法」である。前者は、空集合から始めて、検討対象を次第に広げる。大魚を逸したり偽物にまどわされる危険があるが、普通は効率が高い。かならずしも情報同族体の全部を知る必要がないときによい。後者は全体を情報同族体候補として囲いこみ、次第にせばめていく。はがゆく思われることも多いかも知れないが、情報同族体およびまぎらわしいもの全部をつねに囲いこんでいるところが特徴である。

「あみ法」は構造解析にびったりあう。したがって基礎を「あみ法」におき、必要に応じて「つり法」を併用するのが得策である。

問題は情報同族体の確定のしかたである。対象と情報が固定しても、情報同族体は実は確定しない。それが確定するのは化学が完了する日である。本研究での解析基準としていわゆる(1)頻度の高いものだけに注目する標準用、(2)頻度が低いものをも考慮する拡大用、(3)利用者が自作の基準を組こめる特殊用を設定し、柔軟な対応を目指す。

[システム構築]上の主張の実現性を検定するためにプロトタイプシステムを試作した。文献から材料をとり、原著者らによる結果との比較を行なった。結果の一部を紹介する(図2)。本研究の目から見るとだいたんに絞りこむ両方の原著者は「つり法」志向のように見えるが、その解析結果はおそらく正しい。

例 1 情報源 :

田中誠之 編著、「演習・有機化合物構造決定法」、産業図書、
1983,東京、問題14 (p. 28/スペクトル、pp. 62-63/解答)

情報 : 各種スペクトル および 原著者らによる解析知見

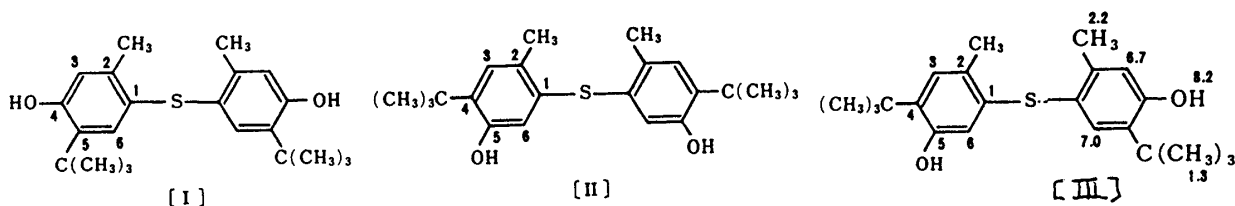
回答

同書

本研究

[I], [II] の2件

[I], [II], [III] など50件



例 2 情報源 :

kimito Funatsu, C. A. Del Carpio. Shin-ichi Sasaki,
Computer Enhanced Spectroscopy, 3, 133 (1986)

情報 : 分子式、スペクトル2種、 および
[ACH 4個、AC 2個、-CH=, -O-, -N=各1個] (A:aromatic)

回答

同論文

本研究

部分構造3個 (I, II, III) (重複あり)

1 など8個

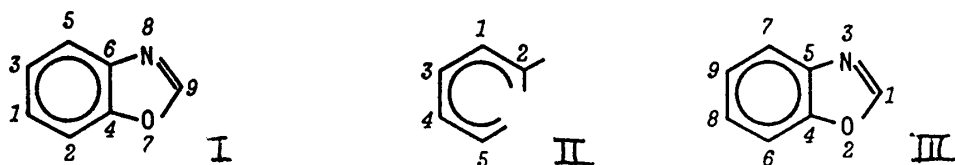


図2 「あみ法」実行結果と普通の方法との比較