

外平面的化学グラフの立体異性体に対する 構造表現および生成アルゴリズム

今田 友樹

(京都大学工学部情報学科 現所属・同大学院情報学研究科数理工学専攻)

指導教員 永持 仁 教授

1. 序論

化合物の立体異性体の列挙は化合物の構造解析や設計に役立つと考えられている。また化合物の多くは外平面グラフの構造をもつことが知られている。同じ結合関係を持つ二つの化合物を立体配置の違いにより異性体として区別する仕方には様々なものが知られているが、本研究では立体中心 (stereocenter) の組合せの違いによる立体異性体に関して、与えられた外平面的化学グラフの立体異性体を列挙する効率的なアルゴリズムを開発することを目的とする。

m 個の立体中心を持つ化学グラフに対しては 2^m 個の立体構造の組み合わせが考えられるが、そのうちのいくつかは全体として同じ立体異性体となっている可能性がある。そこで Nourse[2]はこの 2^m 通りの組み合わせをグラフの対称群によりいくつかの同値類に分割し、その代表元のみを生成することで立体異性体を漏れなく重複なく列挙するアルゴリズムを提案した。このアルゴリズムに対してこれを拡張したものおよび生成ソルバの設計等が行われてきたが、 $O(2^m)$ 時間より効率のよいものは知られていない。最近、大田[3]は外平面的化学グラフに対して動的計画法に基づいてより効率的な立体異性体の列挙アルゴリズムの枠組みを与えた。本論文では大田の与えたアルゴリズムの枠組みに基づき、立体中心に基づく立体異性体の数学的な定義および立体構造の表現方法を新たに定式化し、実際にアルゴリズムを実装し、高速な計算を実現するための諸問題を解決した。

2. 問題および用語等の定義

本研究で扱う問題は次のように定式化される。

入力 C, H, O, N から成る化学物質とそのグラフ構造。各原子には $i=1, 2, \dots, n$ という番号が振られているとする。

出力 立体中心に基づくすべての立体異性体。

ただし出力の立体異性体の正確な定義については本節の最後で述べる。なお入力の化学グラフを G とおき、番号 i の振られた原子に対応する節点を v_i とおく。

本節ではまず立体異性体の構造を表現する節点ラベルの付け方を定式化した。節点 v に対応するラベルを $l(v)$ で表す。この節点ラベルを用いて G の任意の立体異性体の構造表現 I を次のような節点番号とラベルの組の集合で定める。

$$I = \{(i, l(v_i)) \mid i \in \{1, 2, \dots, n\}, (l(v_i) \in \{+, -\})\}$$

この構造表現に対して次の定義に対応する signature を σ_s として定義した。

定義1 G に対する二つの立体異性体表現 I, I' に対して $\sigma_s(I) = \sigma_s(I')$ であるときおよびそのときに限り I と I' は同じ立体異性体を表現する。

この σ_s によって本研究で扱う立体異性体が定義される。つまり、問題の定式化の出力の「すべての立体異性体」とは「 σ_s の異なるすべての立体異性体に対する構造表現 I 」として定義される。

3. アルゴリズム

アルゴリズムの概要は以下のようにになっている。

与えられた G に対して、その構造的中心 (center, centroid) を Jordan の定理[1]を用いて定めることができる。アルゴリズムではまず G の構造的中心を求め、これを根とする。

次に根から遠いところからグラフを上にとりながら各点 v に対して v を根とする根付きグラフの立体構造の数を計算していき、根に到達したときに化学グラフの立体異性体の数を求める。このようにして立体異性体の数 $f^*(G)$ のみを先に計算する。これを異性体の数え上げフェーズと呼ぶ。

さらに数え上げフェーズとは逆に根からグラフを下にとりながら各点 v に対して v のラベルおよび v の子 x, y, \dots の立体構造の番号 k_x, k_y, \dots を定める。このようにして k 番目の立体異性体を列挙し、これ

表1 実験結果. $|f^*(G)|$ が実験により得られた立体異性体の数. t はすべての立体異性体を出力するのにかった時間.

入力	$ f^*(G) $	t (秒)
グラフ構造 1	4,194,304	90.20
グラフ構造 2	2,098,176	43.91
グラフ構造 3	2,097,152	42.92
グラフ構造 4	2,098,176	43.03
グラフ構造 5	2,097,152	41.59
グラフ構造 6	524,800	10.96
グラフ構造 7	1,048,576	21.31
グラフ構造 8	2,097,152	43.11
グラフ構造 9	349,568	12.85

を $k=1, 2, \dots, f^*(G)$ に対して行うことですべての立体異性体を列挙する. これを異性体の構築フェーズと呼ぶ.

4. 計算時間の考察

設計したアルゴリズムに対してその計算時間は次のようにして評価されることを示した.

定理 2 節点数 n の木構造化学グラフ $G=(V, E)$ に対して, 異性体の数え上げフェーズは $O(n)$ 時間で動く. また異性体の構築フェーズは G の立体異性体一つあたり $O(n)$ 時間で動く.

定理 3 節点数 n の外平面的化学グラフ $G=(V, E)$ に対して, 異性体の数え上げフェーズは $O(n)$ 時間で動く. また異性体の構築フェーズは G の立体異性体一つあたり $O(n^3)$ 時間で動き, G の構造的な中心が回転対称性を持つ二連結成分でないときには立体異性体一つあたり $O(n)$ 時間で動く.

5. 実験結果および考察

設計したアルゴリズムを実装して実験を行った. 実験対象として先行研究[4]で $k=1, 2, 3$ について異性体の数が計算済みであるイノシトールおよびその k 個の鎖状縮合化合物を用いた. $k \geq 3$ のときはグラフ構

造の異なる化合物がいくつか存在するので, これらを区別して入力とした.

実験は Intel Pentium 4 CPU 3.00 GHz 搭載の PC で行った. $k=1, 2, 3$ については 1 秒以下の計算時間で[4]の結果と同じ数の立体異性体を出力した. $k=4$ の実験結果を表 1 に与える. 立体異性体の数が少ないほど計算時間も短くなっており, 必要な部分だけを取り出して効率よく列挙が行えていることが分かる.

6. まとめと今後の課題

本論文では立体異性体の構造を表現するラベルの付け方を定式化した上で, 大田[3]の構築した枠組を元にして外平面的化学グラフの立体異性体を列挙する効率的なアルゴリズムを設計した. さらにアルゴリズムを実装して簡単な実験も行った.

今後の課題としては, より多くのタイプの立体異性体を考慮に入れたアルゴリズムの構築や, より広いクラスのグラフ構造を扱えるアルゴリズムの構築等が挙げられる.

参考文献

- [1] C. Jordan, Sur Les Assemblages De Lignes, *J. Reine Angew. Math.*, Vol. 70, 185-190, 1869.
- [2] J.G. Nourse, The configuration symmetry group and its application to stereoisomer generation, specification, and enumeration, *Journal of the American Chemical Society*, Vol. 101, No. 5, 1210-1216, 1979.
- [3] S. Ota, An algorithm for generating stereoisomers of outerplanar chemical graphs, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, 2008, Master's thesis.
- [4] C. Rücker, R. Gugisch and A. Kerber, Manual construction and mathematics- and computer-aided counting of stereoisomers, the example of oligoinositols, *Journal of Chemical Information and Computer Sciences*, Vol. 44, 1654-1665, 2004.