

# 機械学習による 反応経路の予測と生成

岡野原 大輔

化学反応の経路（反応機構や中間状態の系列）を解明・予測することは、新材料の開発やプロセスの最適化に不可欠であり、きわめて重要な課題である。化学反応に関与する原子の配置（原子座標）を入力として、それぞれの構造に対応するエネルギーを地形の「高さ」と見なすと、反応経路とは**反応物**（IS=Initial State）から**生成物**（FS=Final State）へと至る、できるだけエネルギーの低い経路、すなわち**最小エネルギー経路**（MEP=Minimum Energy Path）をたどるものとして理解できる（図1）。この経路上に存在するエネルギーの山越えポイントが「峠」に相当し、**遷移状態**（TS=Transition State）とよばれる。

反応物から遷移状態までのエネルギー差が活性化エネルギーである。また、数学的にみると遷移状態はポテンシャルエネルギー関数の鞍点<sup>あんてん</sup>に対応し、反応経路探索は、最も活性化エネルギーの低い遷移状態（鞍点）をたどっていく経路を探す問題といえる。

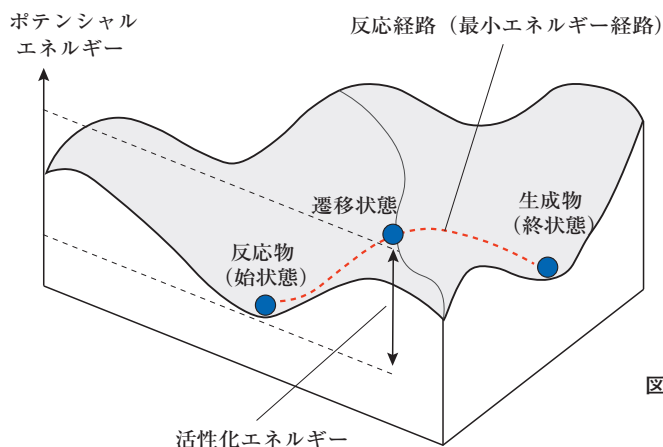
## 化学反応はレアイベントである

こうした化学反応は、分子動力学（MD）シミュレーションを用いて原理的には探索可能である。このような反応を扱うためには結合状態の変化を動的に記述できる反応力場や、ニューラルネットワークポテンシャルの導入が有効である。しかし、活性化エネルギーが高い反応においては、反応が起こる確率がきわめて低く、通常のMDでは反応を観測することが非常に困難である。これを定量的に理解するために、アレニウスの式を用いる。

$$k = A \exp\left(-\frac{E_a}{k_B T}\right)$$

ここで、 $A$ は頻度因子（通常 $10^{12} \sim 10^{13} \text{ s}^{-1}$ ）、 $E_a$ は活性化エネルギー（eV）、 $T$ は絶対温度（K）、 $k_B$ はボルツマン定数である。

反応の平均発生時間 $\tau$ は、反応速度定数 $k$ の逆数で近似される。



反応物から生成物まで、できるだけエネルギーが低い経路をたどっていくのが反応経路である。遷移状態と反応物とのエネルギー差が活性化エネルギーである。

図1 反応経路推定 この図では2次元座標上のポテンシャルエネルギー関数を描いてあるが、実際は（3×原子数）次元上のポテンシャルエネルギー関数である。

$$\tau \simeq \frac{1}{k}$$

たとえば室温（300 K）において、活性化エネルギーが0.3 eV（約 29 kJ mol<sup>-1</sup>）の場合、平均発生時間は約10ナノ秒（10<sup>-8</sup>秒）である。一方、1 eV（約 96 kJ mol<sup>-1</sup>）の場合、平均発生時間は約2時間に達する。

0.3 eV程度の反応は室温でも進行しやすい反応で、たとえば水素分子の金属表面への吸着や固体中の原子の自己拡散などが該当する。一方、1 eV程度の反応は、たとえばC-H結合の切断のように進行しにくい反応である。ここで、シミュレーションの1ステップが1フェムト秒（水素原子間の結合振動が10フェムト秒）とすると、0.3 eVの反応が起こるまでには約10<sup>7</sup>ステップ、1 eVの場合は最大で10<sup>18</sup>ステップが必要となる。

このように、化学反応は非常にまれな事象、レアイベントであり、通常のMDでは現実的な計算時間内に反応が起こらないことが多い。この時間スケールのギャップをどのように埋めるかは重要な課題であり、これまで多くの研究がなされてきた。

たとえば、NEB（Nudged Elastic Band）法は初期的な反応経路の推定の中間構造をビーズとして扱い、これらを仮想的なバネで連結し全体のポテンシャルエネルギーを最適化する。また、化学反応を強制的に発生させる手法としては、高温によって反応を加速するTAD（Temperature Accelerated Dynamics, 文献2）や、高圧のパルスを印加することで反応を促進する NanoReactor 法（文献3）などが提案されている。またGRRM（Global Reaction Route Mapping, 文献4）は、ポテンシャルエネルギー面上の極小構造から出発し、遷移状態や解離経路を網羅的に探索す

るフレームワークとして広く使われている。

## 生成モデルを使った 反応経路推定

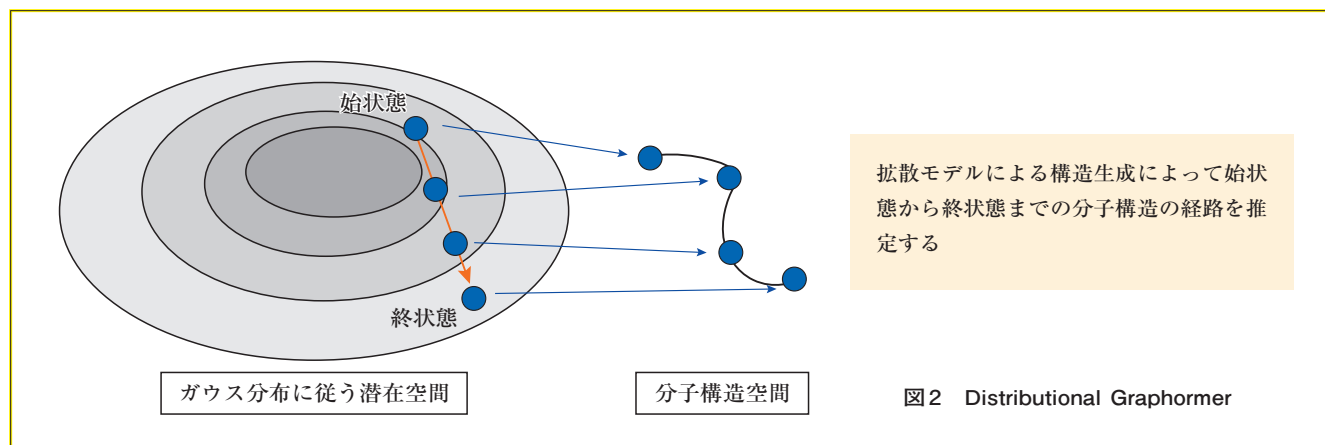
こうした手法を用いても、反応経路を求めることは依然として非常に困難な課題である。正しい反応経路を見つけられなかったり、計算が収束しなかったり、計算コストが非常に高くなるといった問題がしばしば発生する。そうした課題を解決するために、近年では機械学習を用いて反応経路を推定・生成する手法が登場してきている。ここでは、その代表的な手法をいくつか紹介する。

Distributional Graphormer（DiG, 文献5）は、Transformer系のグラフニューラルネットワークと拡散モデルとよばれる生成モデルを組合わせた手法である（図2）。この生成モデルでは、はじめに原子構造の生成を学習し、ガウス分布に従う潜在空間の点と、原子構造が存在するデータ空間の点とが対応づけられる。

次に、始状態（反応物）と終状態（生成物）の原子構造（分子グラフ表現）をそれぞれ潜在空間に埋込み、両者を結ぶ仮想的な直線経路を潜在空間内に定義する。ここでの「直線経路」とは、始状態と終状態の潜在ベクトルを線形補間したベクトル列をさし、潜在空間上で連続的な状態遷移を表現するものである。

さいごに、この補間された潜在ベクトル列の各点から、対応する分子グラフ構造を復元することで、始状態と終状態を結ぶ構造的に滑らかな遷移経路（反応経路）を得ることができる。

このような生成モデルに基づく経路構築手法の利点は、反応メカニズムの詳細をあらかじめ明示的に与えることな



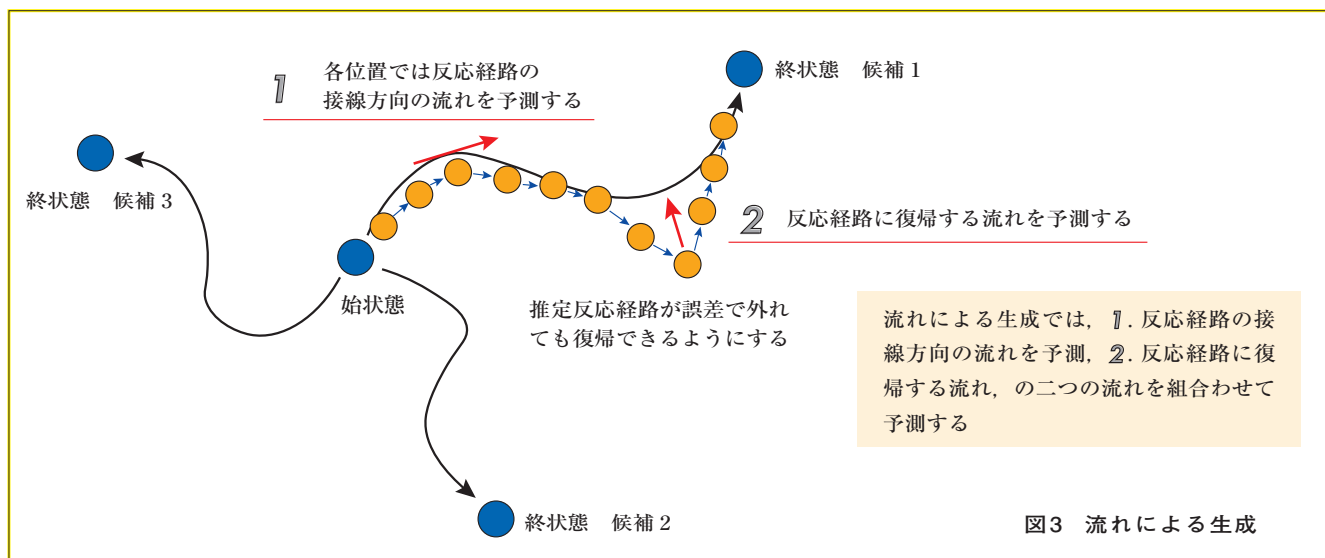


図3 流れによる生成

く、分子構造の情報のみから合理的な中間状態を含む反応経路を得られる点にある。

### 流れに基づく生成モデルによる生成物候補と反応物の同時推定

次に紹介する手法（文献6, 筆者も共著者の一人である）は、始状態のみを入力として与えることで、到達しうる終状態（生成物候補）と、それに至る反応経路を同時に生成するものである（図3）。この手法は、終状態を事前に指定する必要がない点で柔軟性が高く、未知の反応生成物やその反応機構を予測できるという特徴をもつ。

本手法の中心的なアイデアは、反応経路上の構造変化を「流れ（フロー）」を推定する問題として捉えることである。構造の時間発展（または遷移ステップ）において、モデルは以下の2種類の流れをそれぞれ予測するように学習し、それらを統合した流れで反応経路を予測する。

1. 反応経路の接線方向の流れ：現在の構造が自然に変化していく方向、すなわち反応経路に沿った進行方向。

2. 逸脱抑制の流れ：構造が反応経路から逸脱しないように、経路の中心方向に引き戻す補正的な流れ。

この逸脱抑制の流れは、モデルの推論誤差の蓄積や時間発展時の誤差で予測が経路から外れた際に経路上へ戻る

ように設計されており、拡散モデルにおけるデノイジング過程を反応経路に応用している。さらに、反応の種類や制約条件（例：反応させたい原子のペアの指定など）を補助情報として与えることも可能であり、モデルはそれに応じた反応経路と終状態を柔軟に生成する。

実験では、有機反応のデータセットを用いて訓練された本モデルが、訓練に含まれていなかった類似の分子系に対しても、妥当で多様な反応経路を再現可能であることを示している。また、生成された終状態は、実在する化学生成物が含まれていることも報告している。

### 強化学習を使った反応経路推定

次に紹介するのは、Kinetic Monte Carlo (KMC) 法に強化学習（試行錯誤を通して、報酬が最大となる行動を選択するよう学習する手法）を統合することで、原子や分子の動的遷移経路をより効率的にサンプリングする手法である（文献7）。この手法では、個々の反応ではなく、より広いスケールでの反応経路全体を対象とした探索を目指している。

従来のKMC法は、ポテンシャルエネルギー面上の局所安定構造（極小点）から出発し、そこから到達可能な遷移状態（鞍点）への跳躍を、反応確率に基づく確率過程としてサンプリングすることで、反応の動的挙動を模擬する。しかしこの方法では、有望な終状態につながるが、活性化エネルギーが高く、遷移確率がきわめて低いイベントを選択することが難しかった。

本アプローチでは、このサンプリング過程に強化学習の方策 (policy) を使うことを導入し、探索の効率化を進める。具体的には、ある局所構造 (状態) からとりうる構造変化の集合を行動空間とみなし、それぞれの行動がどれだけ将来有望かを評価する価値関数を学習していく。強化学習の枠組みの詳細には踏み込まないが、遷移の履歴とそれに続くエネルギー状態の変化に基づいて、最終的に重要な構造変化につながる行動を優先的に選択するように学習が進められる。これにより、従来のランダムな試行に基づくKMC法とは異なる戦略で効率的かつ選択的に遷移経路をサンプリングすることが可能となる。

この手法は、たとえば金属中における水素拡散経路の予測に適用されており、実験で観測された拡散係数と整合的な遷移系列を再現することに成功している。また、低エネルギーでグローバルな構造変化を伴う遷移経路を高効率で抽出できることも実証されている。

今回はさまざまな手法を紹介してきたが、機械学習による反応経路の予測や生成に関する研究は、依然として発展

途上であるといえる。そのおもな理由の一つは、学習や評価に用いるデータセットが今なお不十分であること、もう一つは、実用に耐えるためには非常に高い予測精度が求められることである。

前回までに紹介したようなポテンシャル関数が、より高精度かつ高速に評価できるようになってきたなか、反応経路予測および反応探索の手法の発展が進むことが期待される。

## 参考文献

1. S. Kimほか, *Nat. Commun.*, **15**, 341 (2024).
2. R. J. Zamoraほか, *Annu. Rev. Chem. Biomol. Eng.*, **7**, 87 (2016).
3. L.-P. Wangほか, *Nat. Chem.*, **6**, 1044 (2014).
4. S. Maedaほか, *Phys. Chem. Chem. Phys.*, **15**, 3683 (2013).
5. S. Zhengほか, *Nat. Mach. Intell.*, **6**, 558 (2024).
6. A. Hayashiほか, "Generative Model for Constructing Reaction Path from Initial to Final States", arXiv: 2401.10721
7. H. Tangほか, *Adv. Sci.*, **11**, 2304122 (2024).