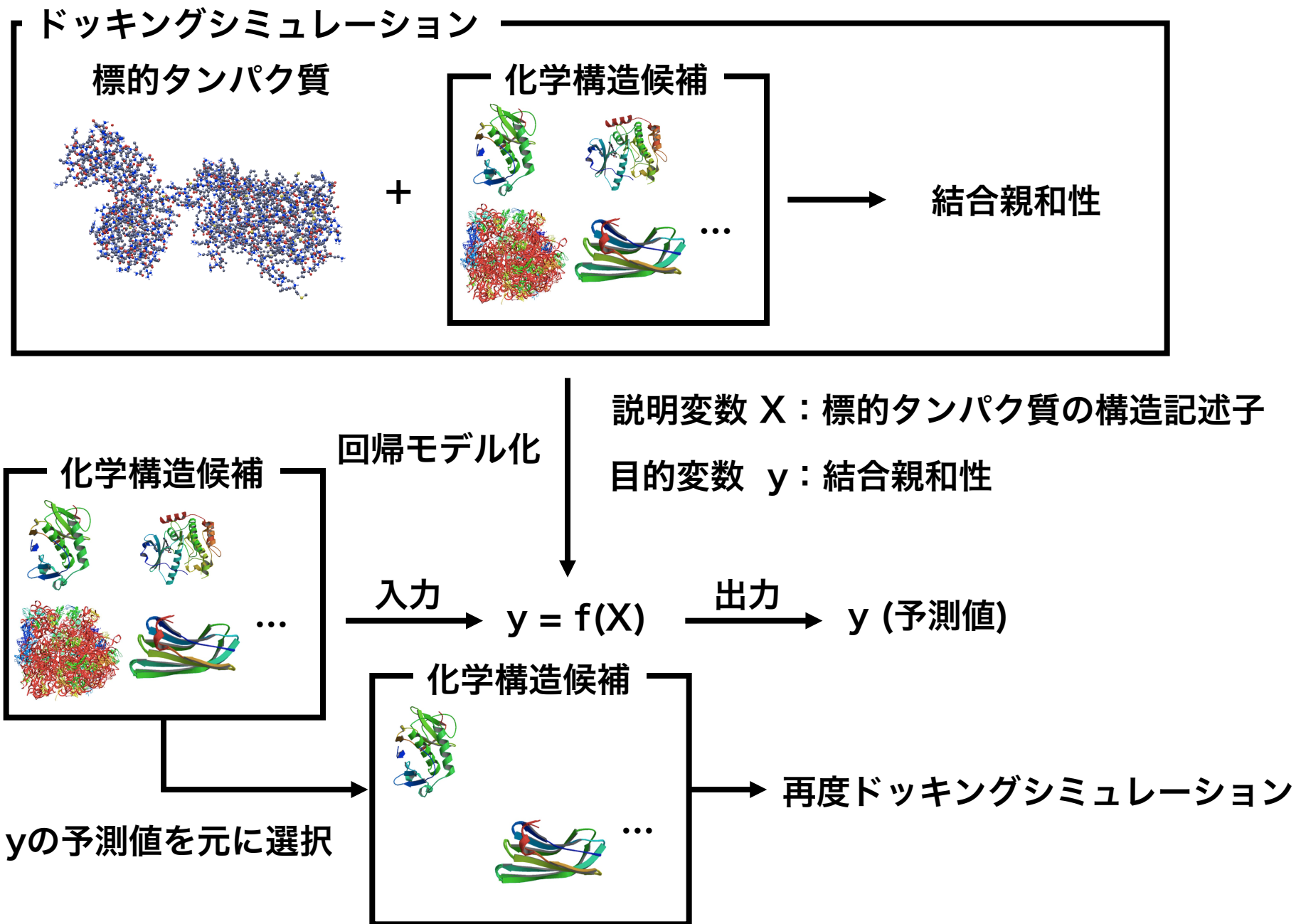


ドッキングシミュレーションおよび 機械学習を活用した医薬品化学構造の設計

May. 12, 2016
ディスカッション

酒井研究室 宮崎 大輝

リード化合物設計



構造記述子の抽出

- 実験計画法

- 一 ある実験結果に対して影響している因子とその水準を変化させデータを得る

(因子：構造記述子 (分子量、ベンゼン環の数など))

因子数の多いシミュレーションに対し、過飽和実験計画法を用いることにより少ない回数で効率よくシミュレーションを行うことが可能

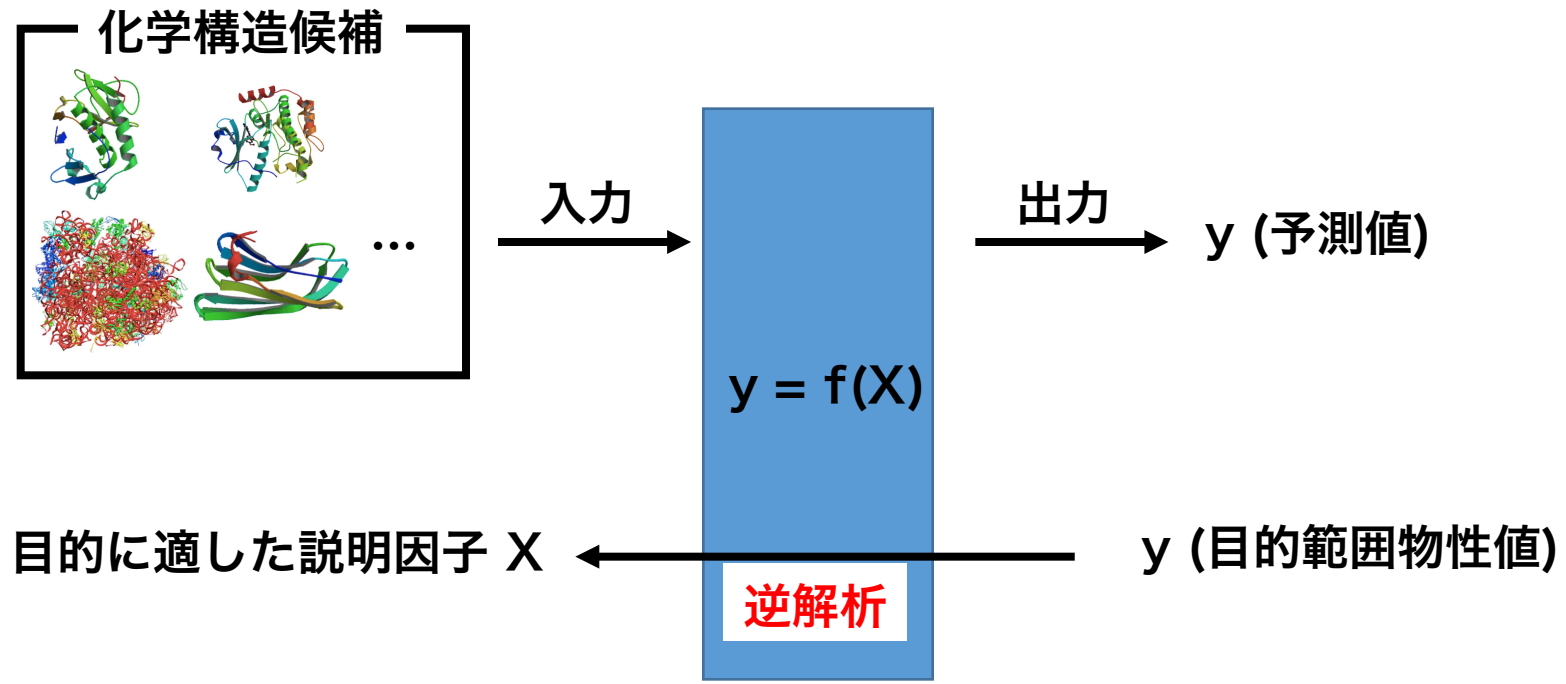
- 寄与率による選定

- 一 回帰モデルにおける各説明変数の寄与率を算出

その寄与率に基づき最適化に用いる因子を決定

構造記述子の抽出

- 逆解析

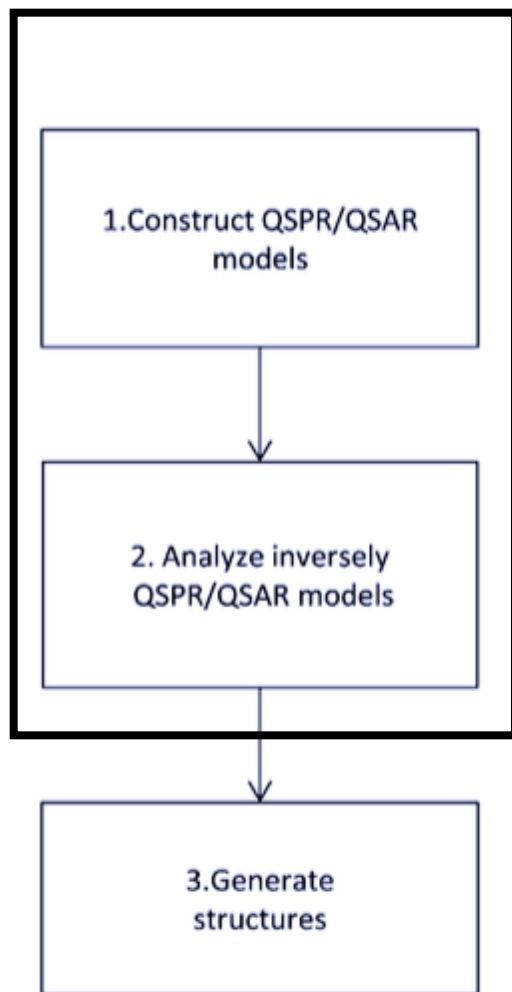


逆解析により得られた構造がモデル式構築時に用いたトレーニングデータとかけ離れた構造である場合、適切な予測ができない可能性

→ 適用範囲 (AD) を適切に定める必要性

ex. 各記述子の範囲、入力サンプルとトレーニングデータとの距離...

逆解析における適用範囲(AD)

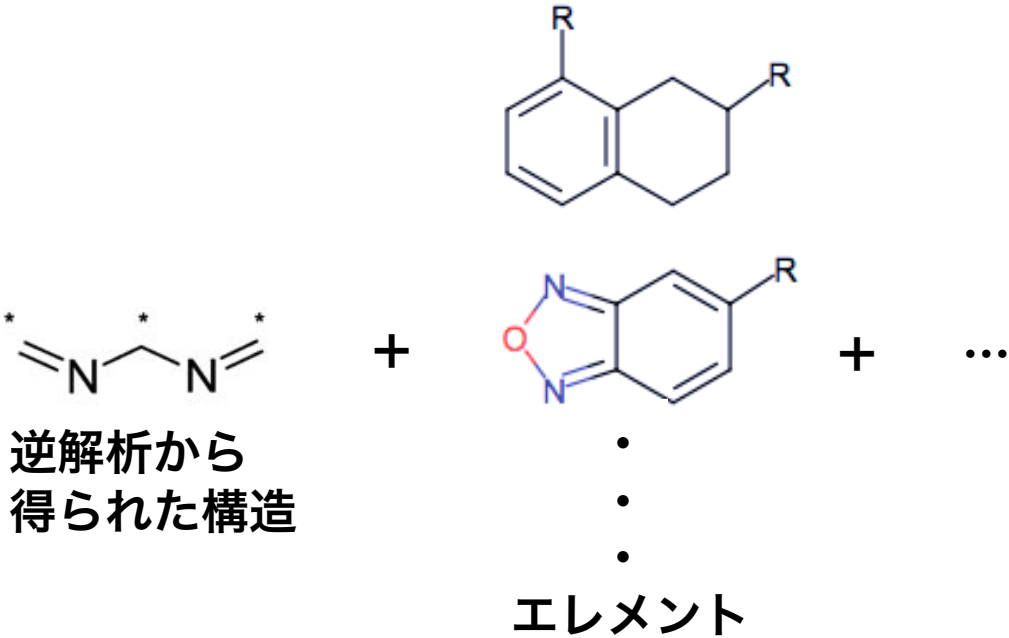
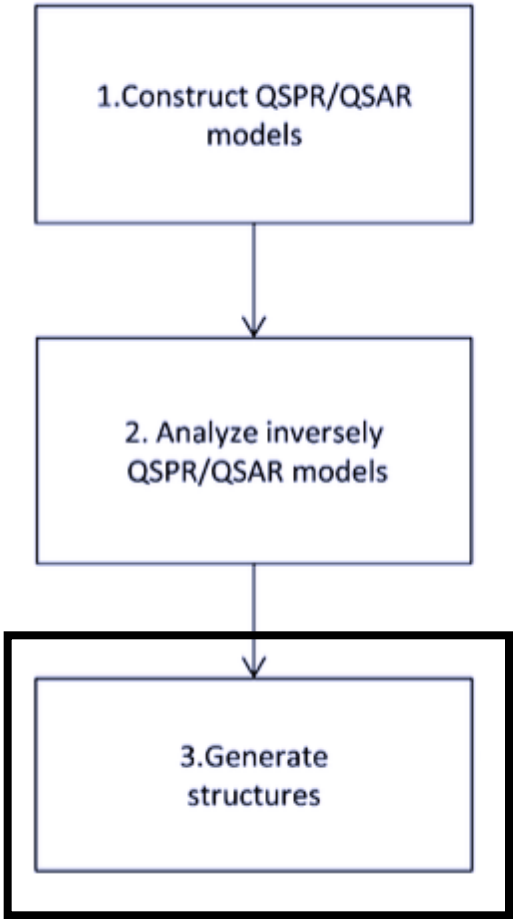


MLRにより得られた条件付き確率 $p(y|X)$ と GMMsにより得られた 周辺確率 $p(X)$ を用い 条件付き確率 $p(X|y)$ を得る。

$p(X|y)$ に従い逆解析を行うことで、 説明変数が確率密度変数として得られる。

→ 予測に適した構造記述子を密度分布として 見積もることができる。

逆解析における適用範囲(AD)



トレーニングデータセット内の分子フラグメントをエレメントとして、Canonical construction path methodによって、重複なく構造を生成する。

→ ADを考慮したのと同様