

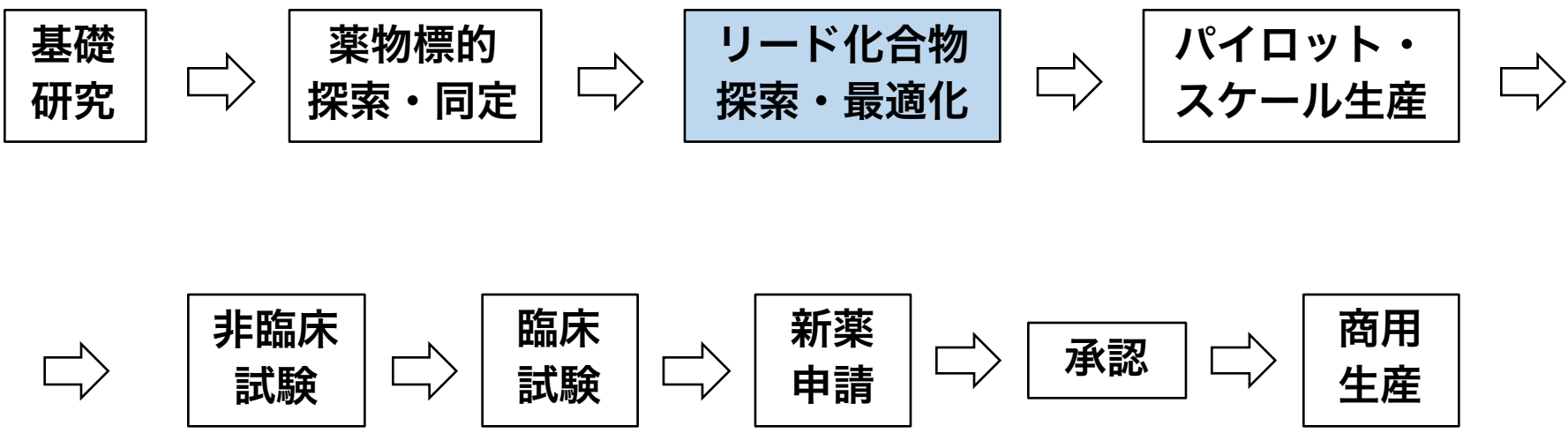
# ドッキングシミュレーションおよび 機械学習を活用した医薬品化学構造の設計

June. 1, 2016  
ディスカッション  
(中間発表練習)

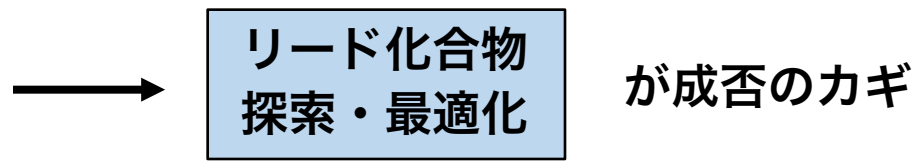
酒井研究室      宮崎 大輝

# 医薬品の開発

## 医薬品開発の一般的な流れ



- ・ 長い年月・莫大な開発費用
  - 開発年数：十数年
  - 成功確率：数万分の一



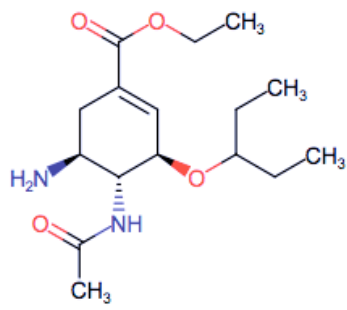
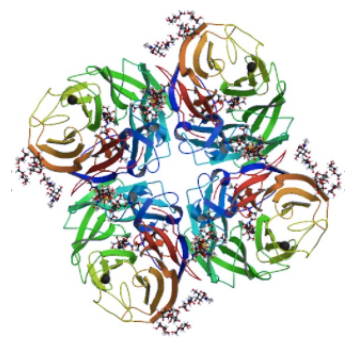
# 医薬品の開発

疾患の原因タンパク質

薬物化合物

インフルエンザ：ノイラミニダーゼ

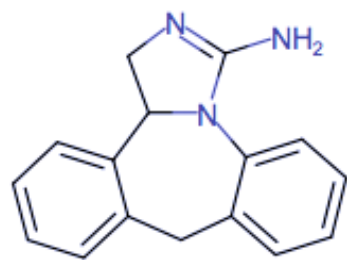
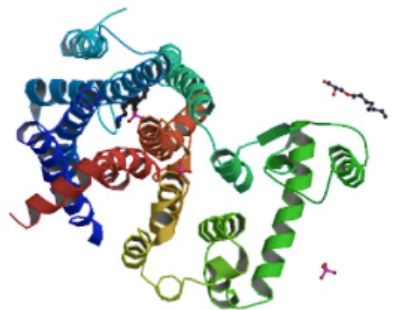
ノイラミニダーゼ阻害剤  
(タミフル・リレンザなど)



感染細胞からのインフルエンザウイルスの放出を妨げる

花粉症：ヒスタミンH1受容体

ヒスタミンH1受容体拮抗薬  
(アレジオンなど)

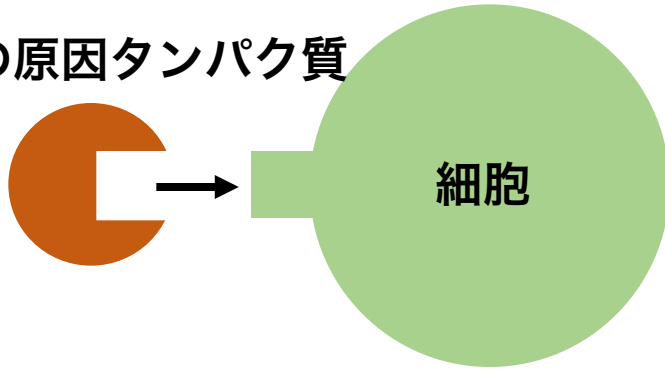


ヒスタミンとヒスタミン受容体との結合を妨げる

→ 薬物の開発へ繋がる出発点となるリード化合物の探索

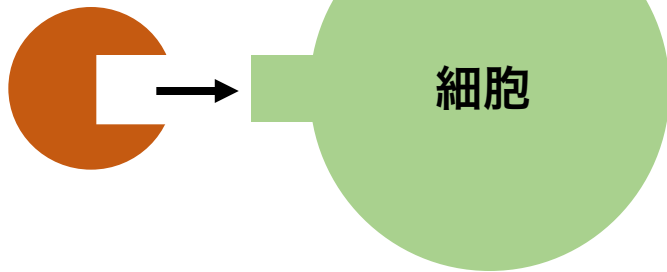
# リード化合物の探索

疾患の原因タンパク質



# リード化合物の探索

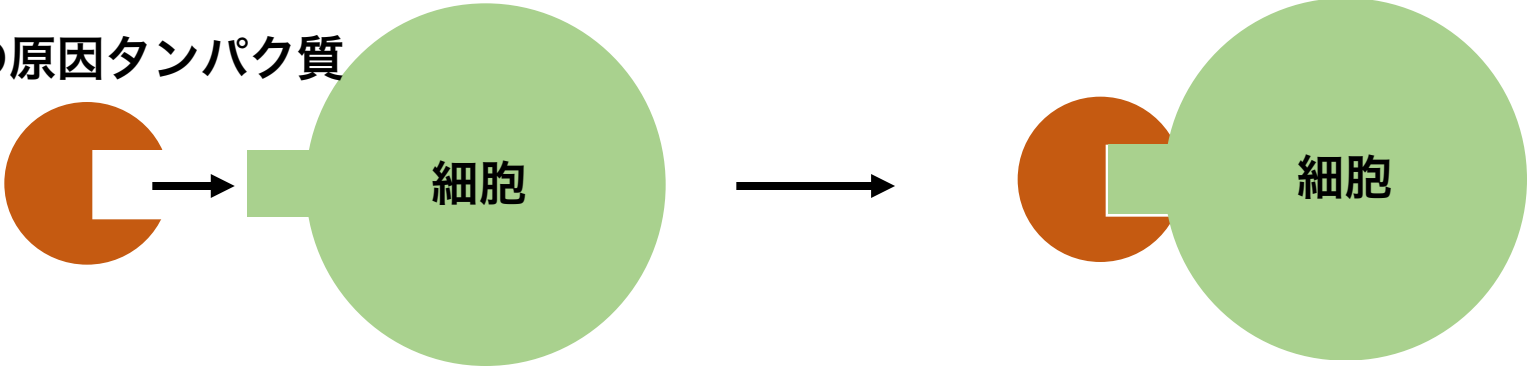
疾患の原因タンパク質



細胞の機能低下など  
→発症

# リード化合物の探索

疾患の原因タンパク質

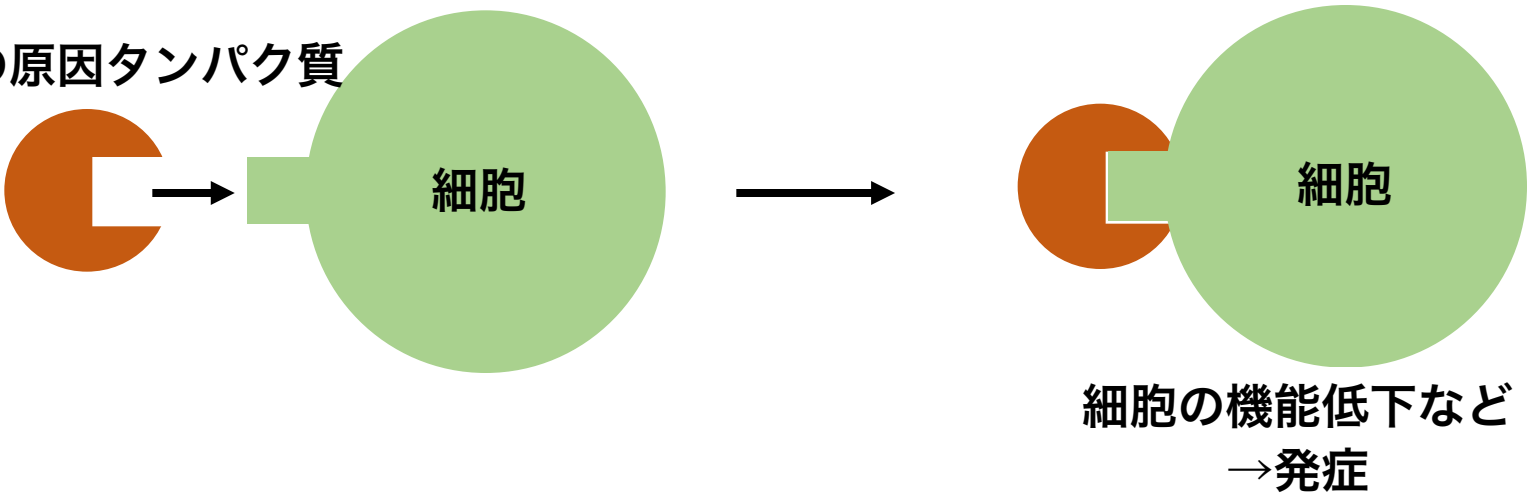


疾患の原因タンパク質    リード化合物

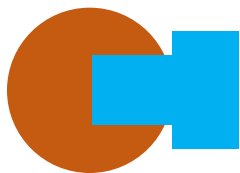


# リード化合物の探索

疾患の原因タンパク質

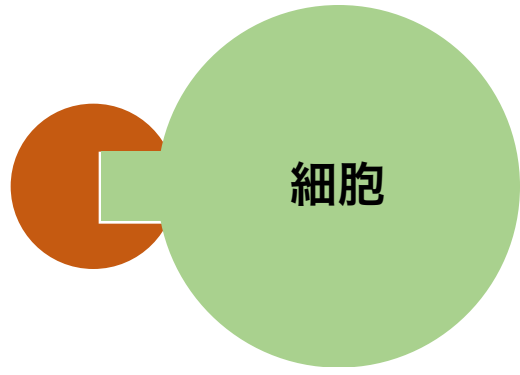
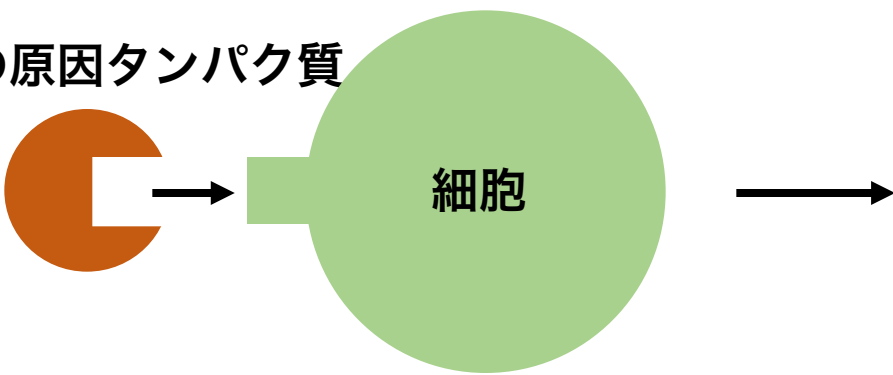


原因分子の不活性化



# リード化合物の探索

疾患の原因タンパク質



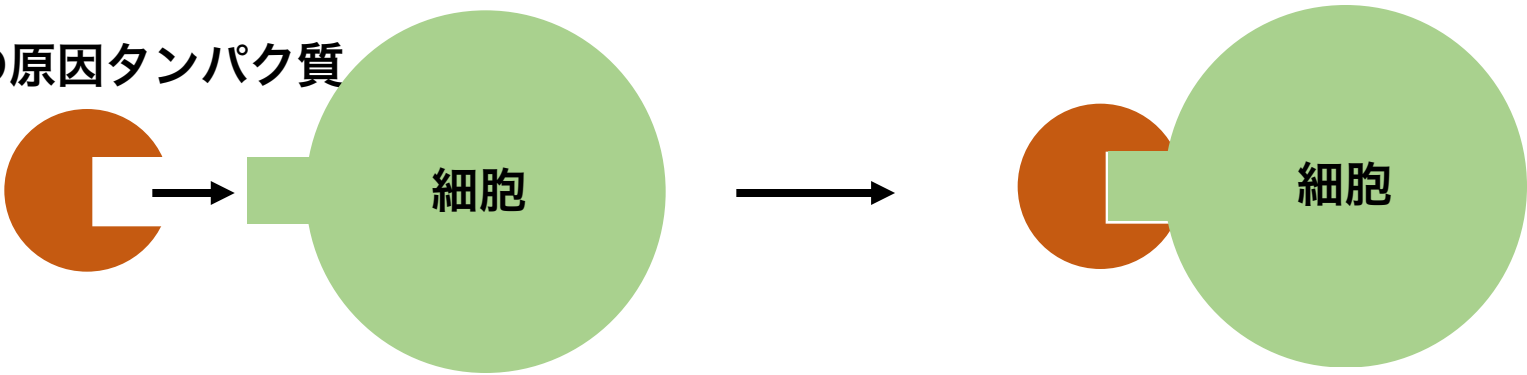
原因分子の不活性化





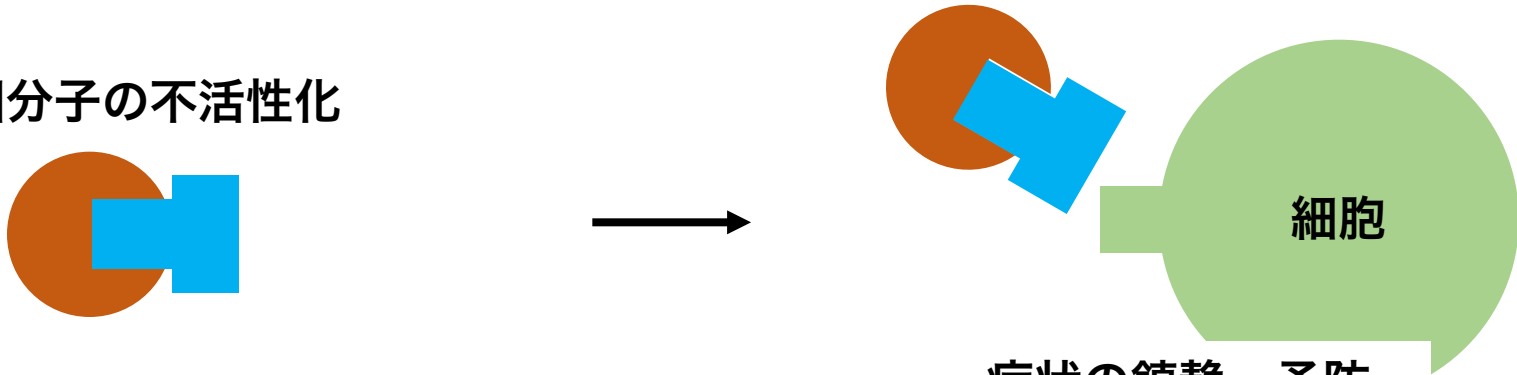
# リード化合物の探索

疾患の原因タンパク質



細胞の機能低下など  
→発症

原因分子の不活性化

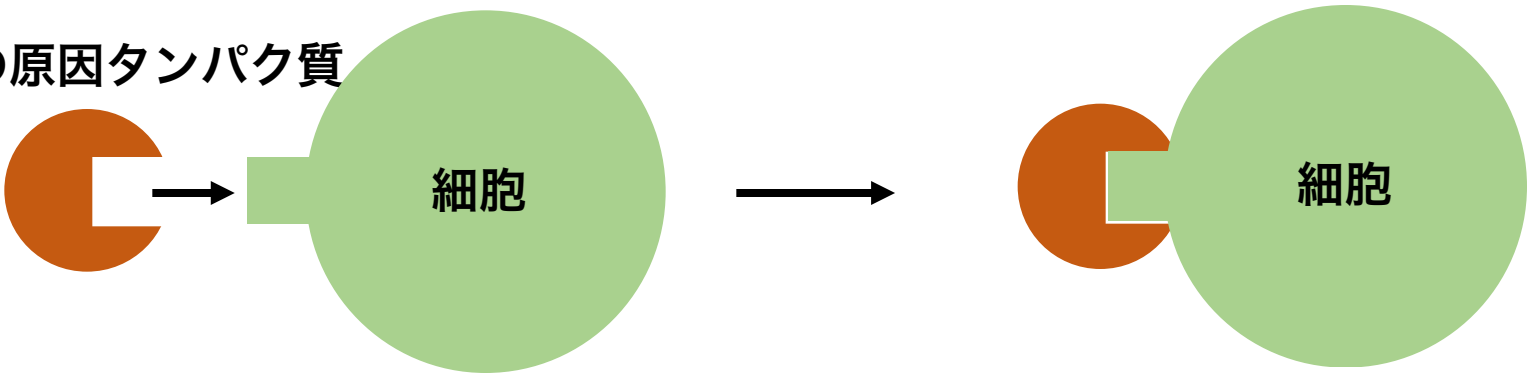


症状の鎮静、予防

————→ 標的タンパク質と化合物との結合構造及び結合親和性予測

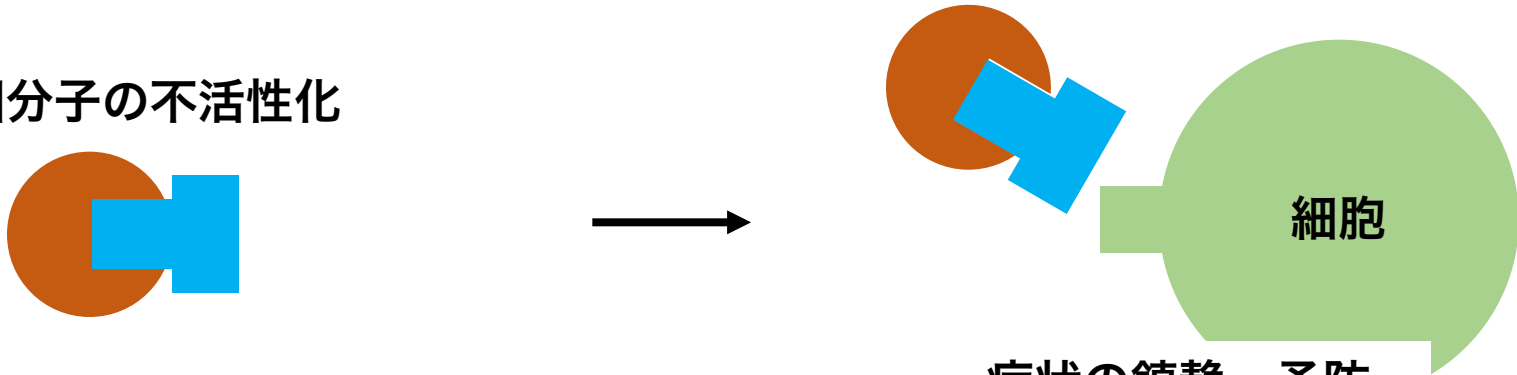
# リード化合物の探索

疾患の原因タンパク質



細胞の機能低下など  
→発症

原因分子の不活性化



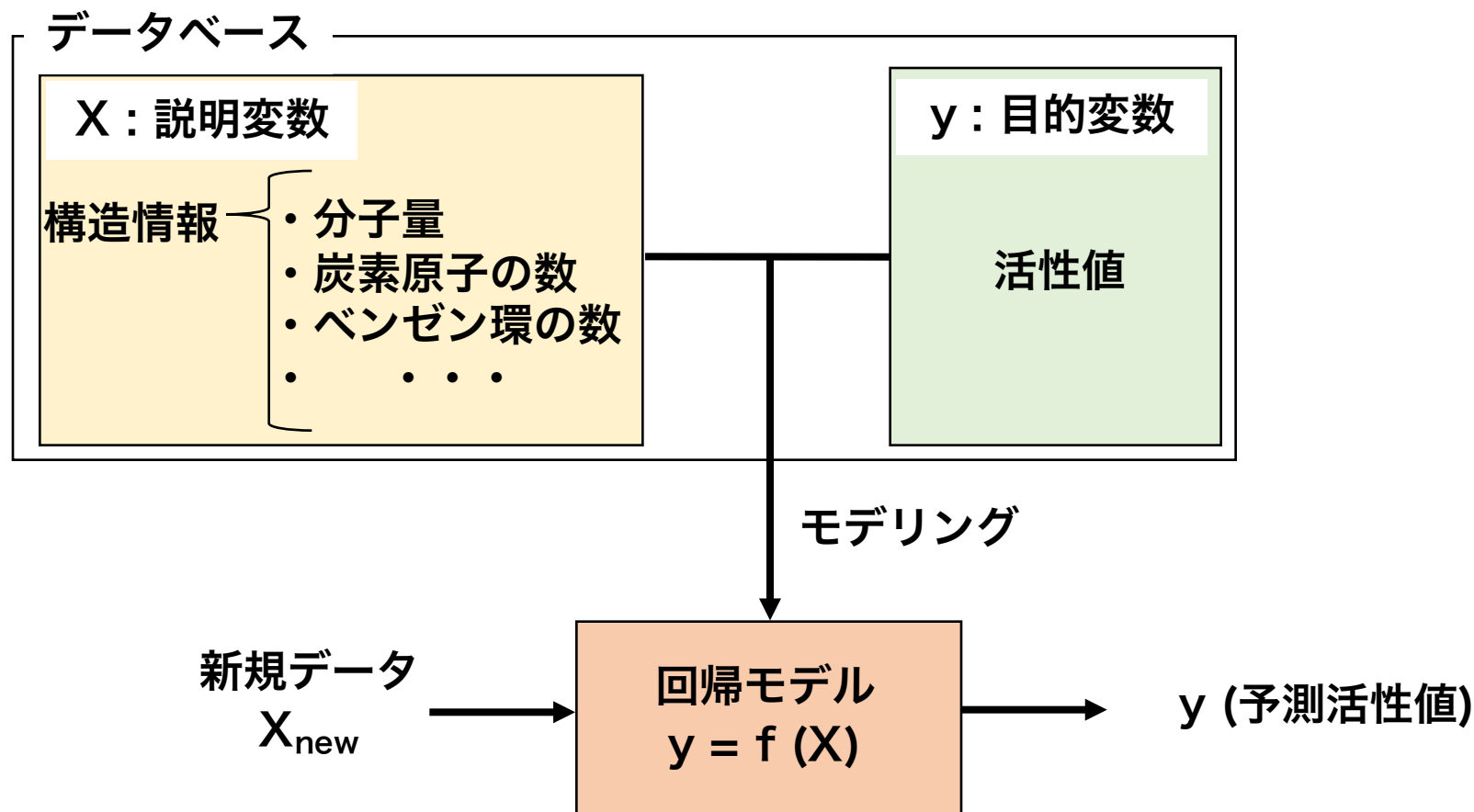
症状の鎮静、予防

- 標的タンパク質と化合物との結合構造及び結合親和性予測
- ドッキングシミュレーション

# リード化合物の探索

## 定量的構造活性相関 (QSAR)

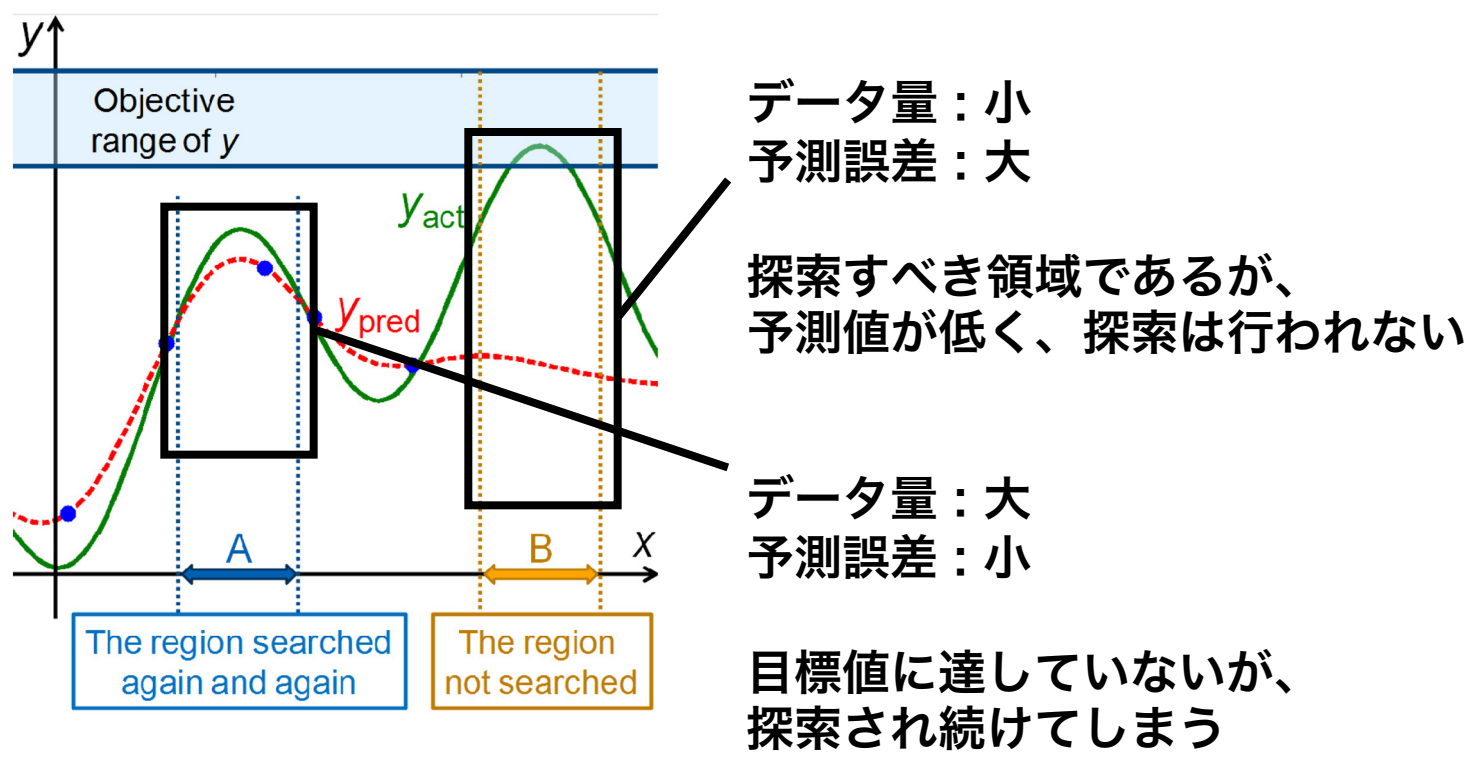
- 既知のデータベースから作成したモデルにより、活性値を予測



→ 機械学習による回帰予測モデルの構築

# 回帰モデルによる探索の問題点

- データ量の大小による予測誤差のために、適切な外挿領域の探索が行われない



- モデル構築に用いた既知データの密度が低いと、予測値の信頼性が低い

→ 予測誤差の大きさとデータ密度の考慮により効率的な探索が可能

# 予測誤差の大きさの推定

## - GP(Gaussian Process)法

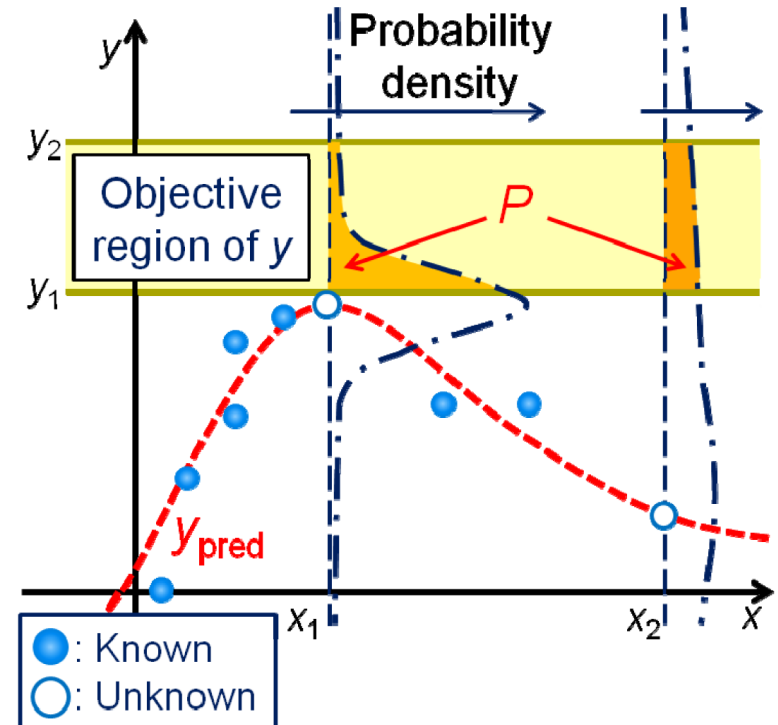
ある説明変数 $x$ が与えられた時に、  
目的変数 $y$ を正規分布に従う確率モデルとする回帰手法

モデル式： $y = w^T \varphi(x)$   
( $\varphi$ ：非線形関数、 $w$ ：回帰パラメタ)

- ・ 予測誤差の分散 $s^2$ を求めることが可能
- ・  $s^2$ を用いて目的物性達成確率 $P$ を算出

$$P = \int_{y_1}^{y_2} \frac{1}{\sqrt{2\pi}s} \exp\left\{-\frac{(y - y_{pred})^2}{2s^2}\right\} dy$$

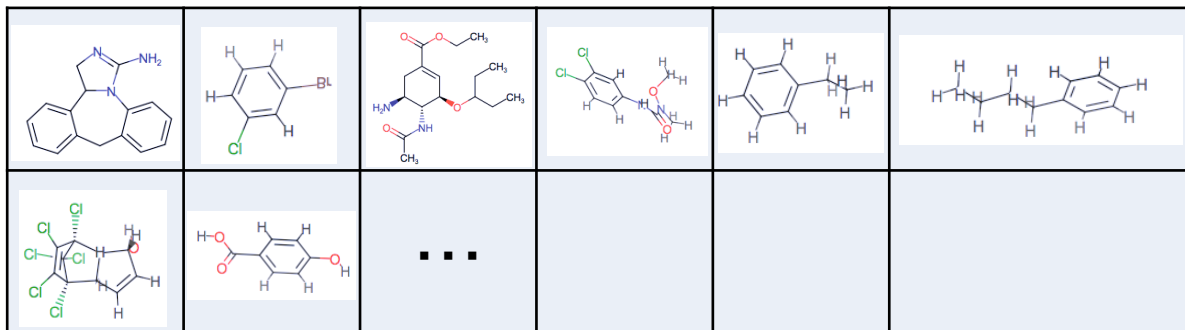
予測誤差の分散が大きい $x_2$ のような候補  
において $P$ が大きい値をとる



→ 探索の効率化

# リード化合物の探索

## 化合物候補データベース



低分子有機構造数  $\gg 10^{60}$  構造

実験及びドッキングシミュレーション  
による全空間探索は不可能

機械学習による  
構造活性相関モデルの精度は不安定

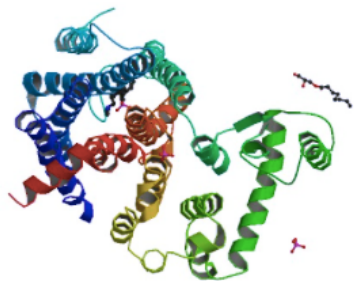


ドッキングシミュレーション  
+  
機械学習による構造活性相関予測

# リード化合物設計

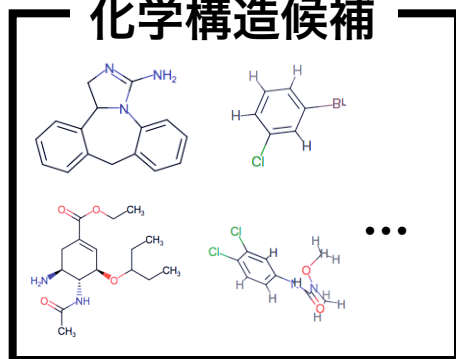
## ドッキングシミュレーション

標的タンパク質



+

化学構造候補



結合親和性

回帰モデル化

説明変数  $X$  : 標的タンパク質の構造記述子

目的変数  $y$  : 結合親和性

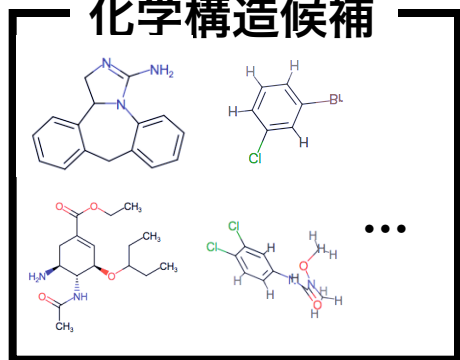
入力

$$y = f(X)$$

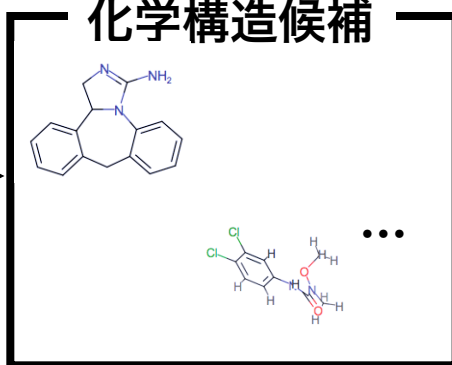
出力

$y$  (予測値)

化学構造候補



化学構造候補



再度ドッキングシミュレーション

$y$ の予測値を元に選択