



# Machine learning designs new GCGR/GLP-1R dual agonists with enhanced biological potency

---

Received: 26 December 2021

---

Accepted: 8 April 2024

---

Published online: 16 May 2024

Anna M. Puszkarska <sup>1,6</sup>, Bruck Taddese <sup>2,7</sup>, Jefferson Revell <sup>2</sup>, Graeme Davies <sup>3</sup>, Joss Field <sup>3</sup>, David C. Hornigold <sup>3</sup>, Andrew Buchanan <sup>4</sup>, Tristan J. Vaughan <sup>4,8</sup> & Lucy J. Colwell <sup>1,5</sup>

B4 中村優作

2024/09/30 後期雑誌会

## 肥満

体脂肪が過剰に蓄積した状態



- 脂肪細胞の増加、肥大
- 肝臓への脂質蓄積



## 肥大した脂肪細胞

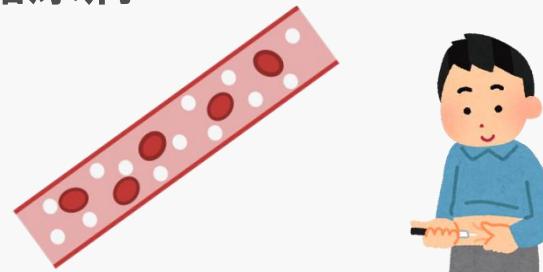
- アディポカインの分泌  
→ インスリンシグナル伝達を直接妨害
- マクロファージの動員  
→ 脂肪細胞のインスリン抵抗性を悪化



## 糖尿病の発症

米国では患者の  
90%以上が肥満

## 糖尿病



高血糖(血中グルコース濃度が高い)  
状態が慢性的に続く症状



インスリンの不足、インスリンに  
対する反応性の低下により、  
血液中のグルコースを組織に  
移動させることができなくなる



網膜症、神経障害、腎臓病、  
末梢血管障害、心臓病、脳卒中

[Micael Wolfe, xPharm, 2007]

## ● GLP-1R (グルカゴン様ペプチド1受容体)

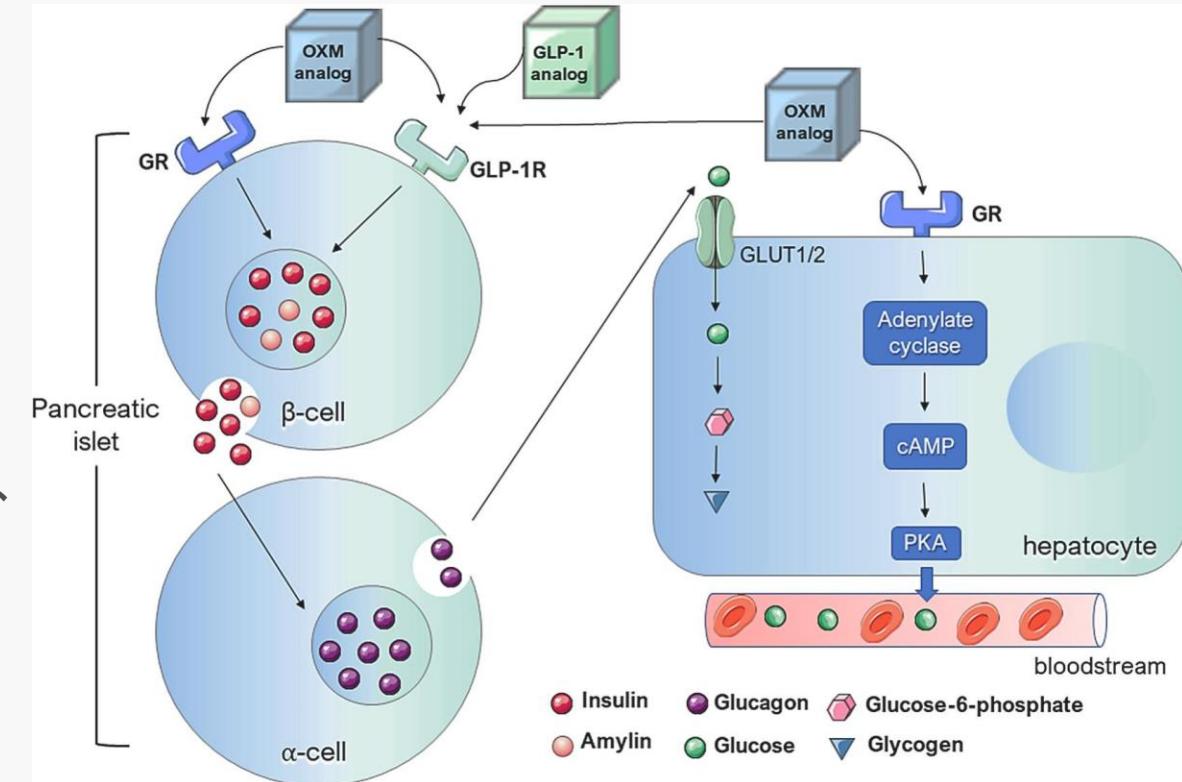
- 脇 $\beta$ 細胞でのインスリン分泌による血糖値低下
  - グルカゴンの分泌制御
  - 食欲低下作用
- 糖尿病治療薬として承認されている

## ● GCGR (グルカゴン受容体)

肝臓での糖産生を増加させ、血糖値を上昇させるが、以下の人体に有益な作用を持つ。

- $\beta$ 酸化による脂肪分解
- 肝臓でのエネルギー消費
- 中枢神経に作用して満腹感の誘導

→ 糖尿病、肥満の治療標的として注目されている



[ Renata Spezami et al., Life Sciences, 2022]

GLP-1R, GCGRは糖尿病、肥満の治療標的

## GLP-1Rの強力な活性化



吐き気などの有害事象  
を誘発する



副作用の少ない、  
より効果的な  
治療法の開発が必要

## GCGR/GLP-1R dual agonist

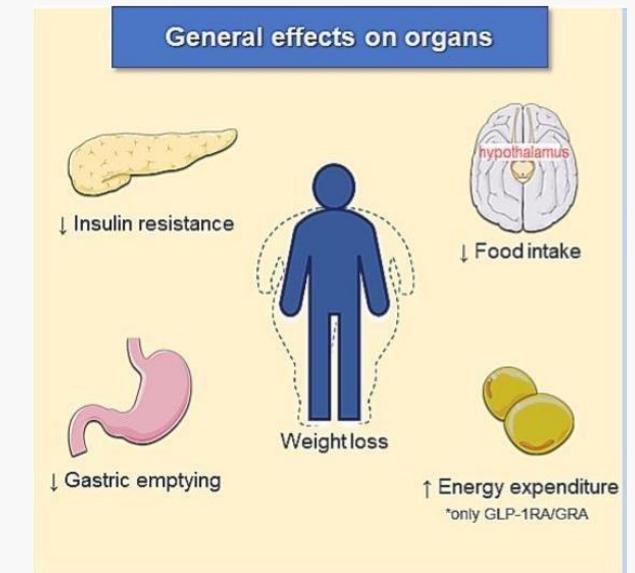
## GCGR

- 満腹感の誘導
- 肝臓での脂肪分解
- ✗ 血糖上昇作用



## GLP-1R

- 血糖値低下作用
- 食欲低下作用
- ✗ (有害な胃腸作用)



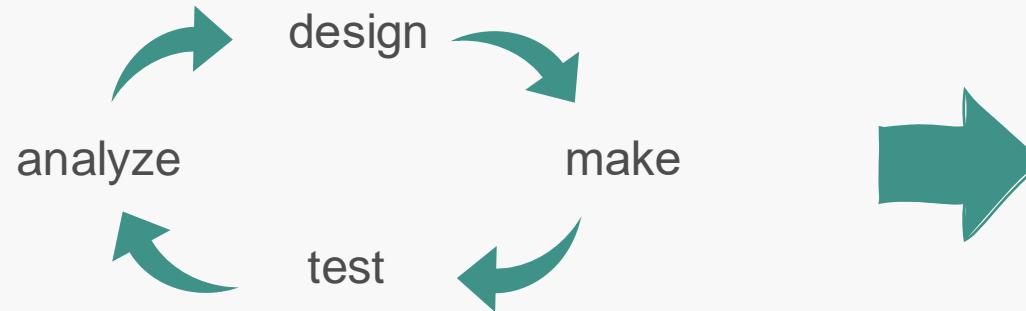
## Dual agonist

- 食欲低下作用
- 満腹感の誘導
- 肝臓での脂肪分解

[ Renata Spezami et al., Life Sciences, 2022]

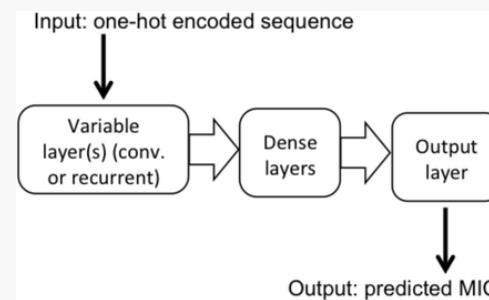
GCGR/GLP-1R dual agonist はより安全な治療薬となる可能性

新しいペプチド設計には時間とコストがかかる



既存の実験データを用いて  
機械学習モデルを学習させることにより、  
時間とコストの解決を目指す

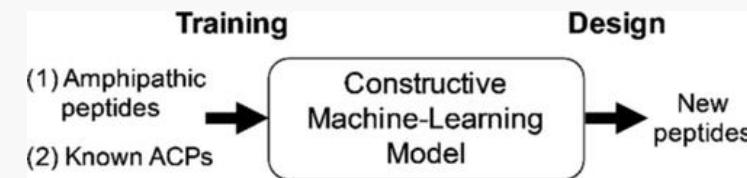
近年のペプチドや低分子の設計のための機械学習モデルの開発の成功



4050件の抗菌ペプチドと  
MIC(最小発育阻止濃度) のデータ

MICが既知のデータの79%  
より低い抗菌ペプチドを  
デザイン

[ Zack Witten et al., bioRxiv, 2019 ]

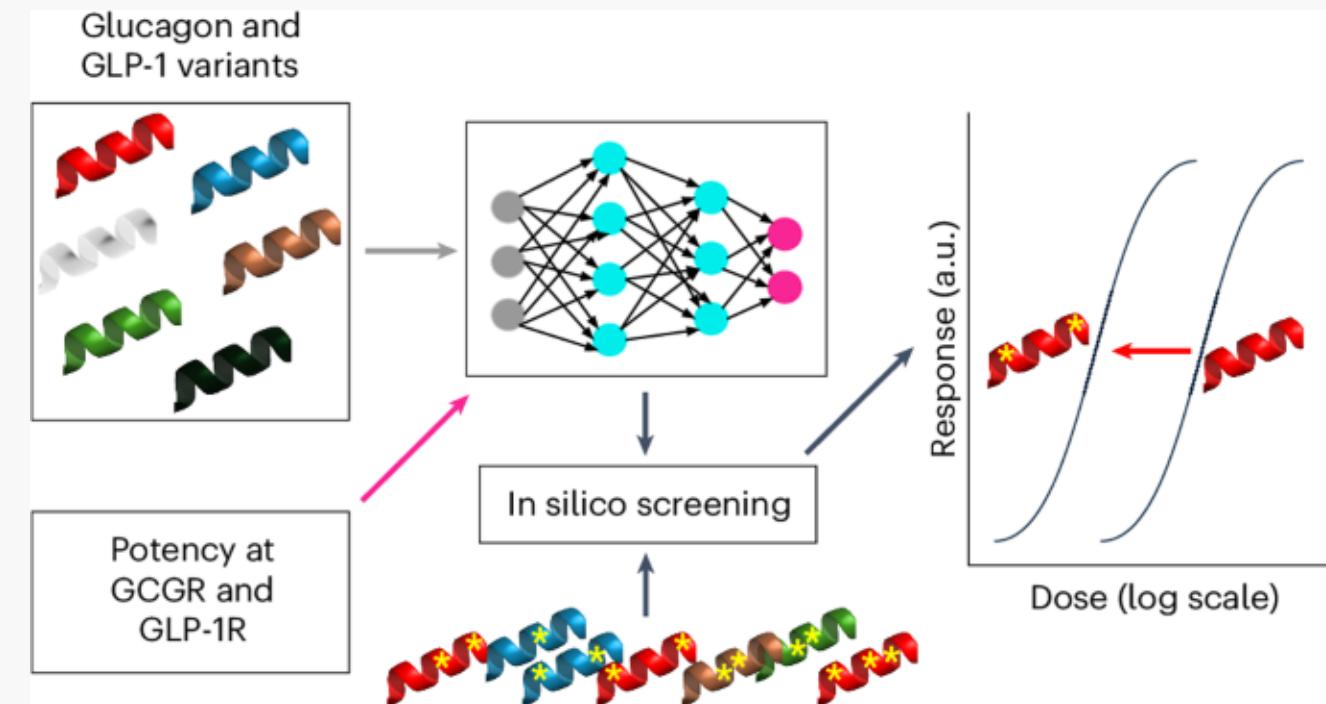


1. 10000個のαヘリカル陽イオン両親媒性ペプチド
2. 26の膜溶解性抗がんペプチドのEC50

12個中10個が *in vitro* で膜溶解性を示した

[ Gisbert Schneider et al., ChemMedChem, 2018 ]

1. 125のGCG及びGLP-1ペプチド類似体のセットを用意し、GCGR及びGLP-1RのEC50を測定
2. GCGRとGLP-1Rの両方において、ペプチド配列からEC50を予測するモデルを訓練
3. データセットから数点残基置換させた、未知のペプチド配列の活性を予測
4. 実際にペプチドを合成し、GCGRとGLP-1Rの活性を確認



## 目的

GCGRとGLP-1Rの両方において、ペプチド配列からそれぞれのEC50を予測するモデルを構築する。

## 新規性

機械学習モデルを用いて、GCGR/GLP-1デュアルアゴニストペプチドの設計に成功した。

## 選定理由

- 自分の研究でTGR5活性化によりGLP-1分泌を促進する目的があり、GLP-1の知識が勉強になると感じたため。
- 機械学習によってデュアルアゴニストペプチドの予測をする手法が面白いと感じたため。

Fig1. トレーニングデータの作成

Fig2. 機械学習モデルの構築と評価

Fig3. 作成したモデルを利用したリガンドのデザイン

Fig4. デザインしたペプチドの合成と評価

Fig5. トレーニングデータとデザインしたペプチドの比較

Fig1. トレーニングデータの作成

Fig2. 機械学習モデルの構築と評価

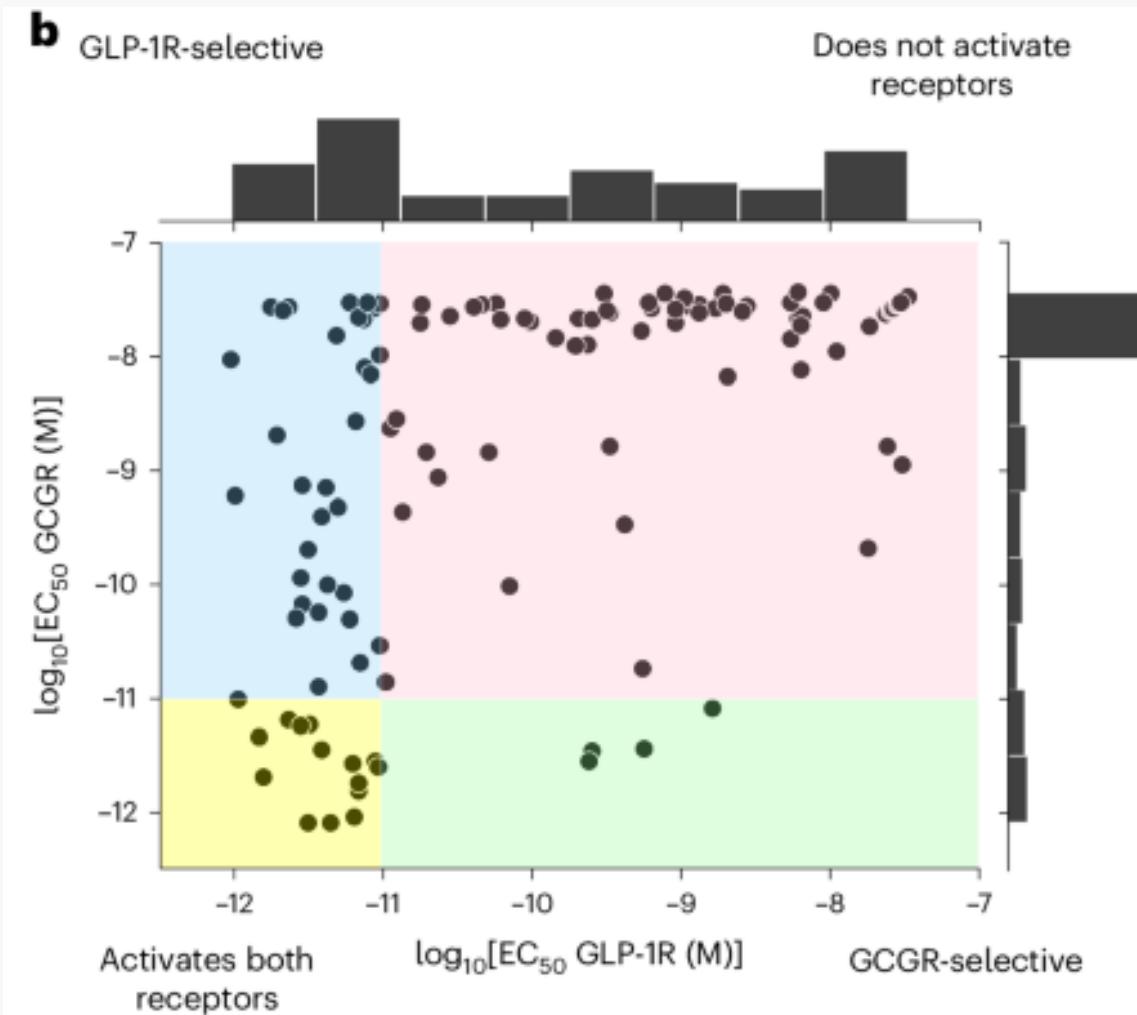
Fig3. 作成したモデルを利用したリガンドのデザイン

Fig4. デザインしたペプチドの合成と評価

Fig5. トレーニングデータとデザインしたペプチドの比較

## 機械学習に用いるデータの測定

- チャイニーズハムスター卵巣(CHO)細胞にhGCGとhGLP-1Rを発現させ、ペプチドを添加してそれぞれのEC<sub>50</sub>値を測定
- 125個のペプチドに対して実験
- データを4つの活性領域に分割
  - 両方の受容体で活性
  - hGLP-1R選択的
  - hGCG選択的
  - 両方の受容体で不活性



125のペプチドのGCG/GLP-1Rに対するEC50のデータを作成

## hGCGとhGLP-1の配列

グルカゴン HSQGTFTSDYSKYLDSRRAQDFVQWLMNT 29残基

GLP-1 HAEGTFTSDVSSYLEGQAAKEFI AWLVKGR 30残基

→ 30残基中14個のアミノ酸が共通

## ペプチド配列のエンコード

機械学習の入力として扱えるようにペプチド配列をエンコード

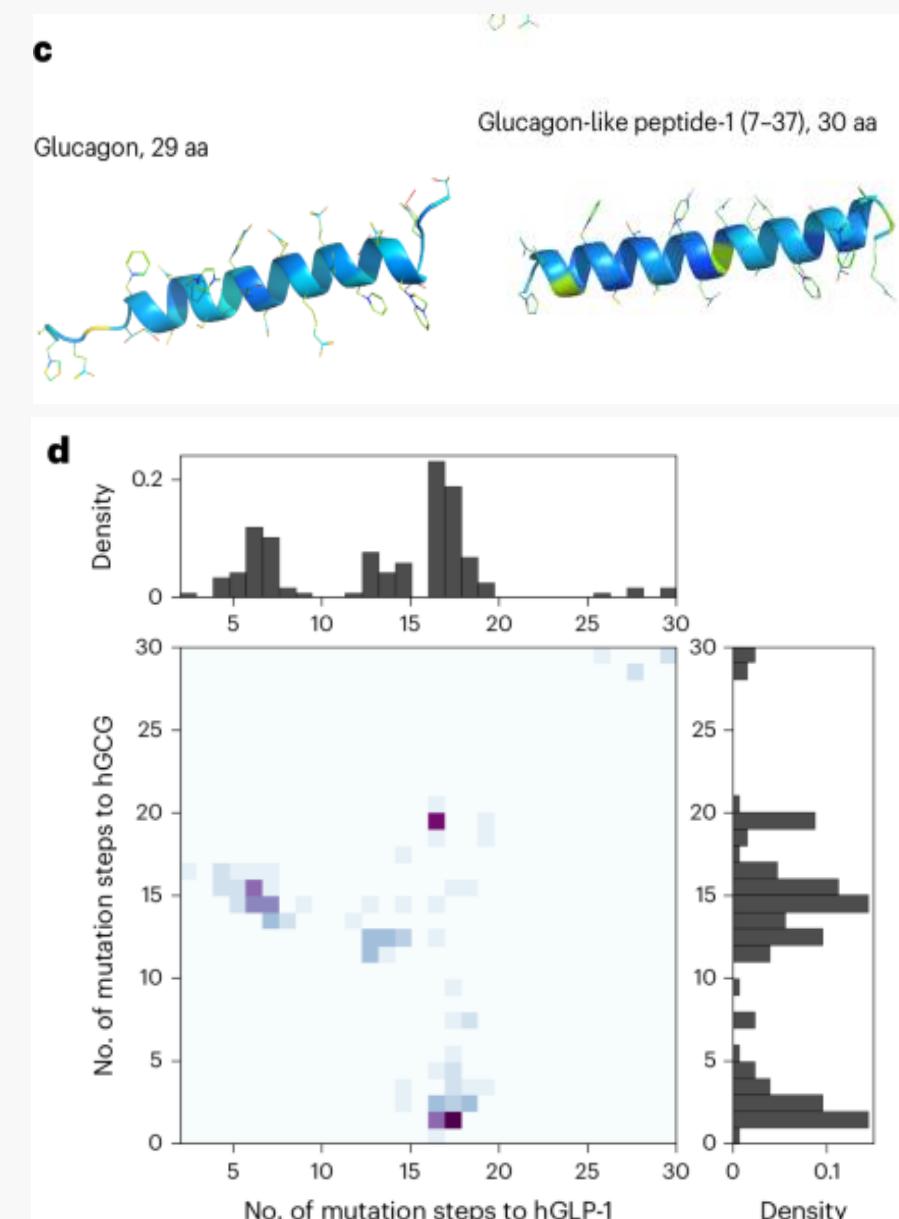
すべてのトレーニングデータを、アミノ酸20と欠損”-”の

計21カテゴリを含む、長さ30の一次元ベクトルとして表現

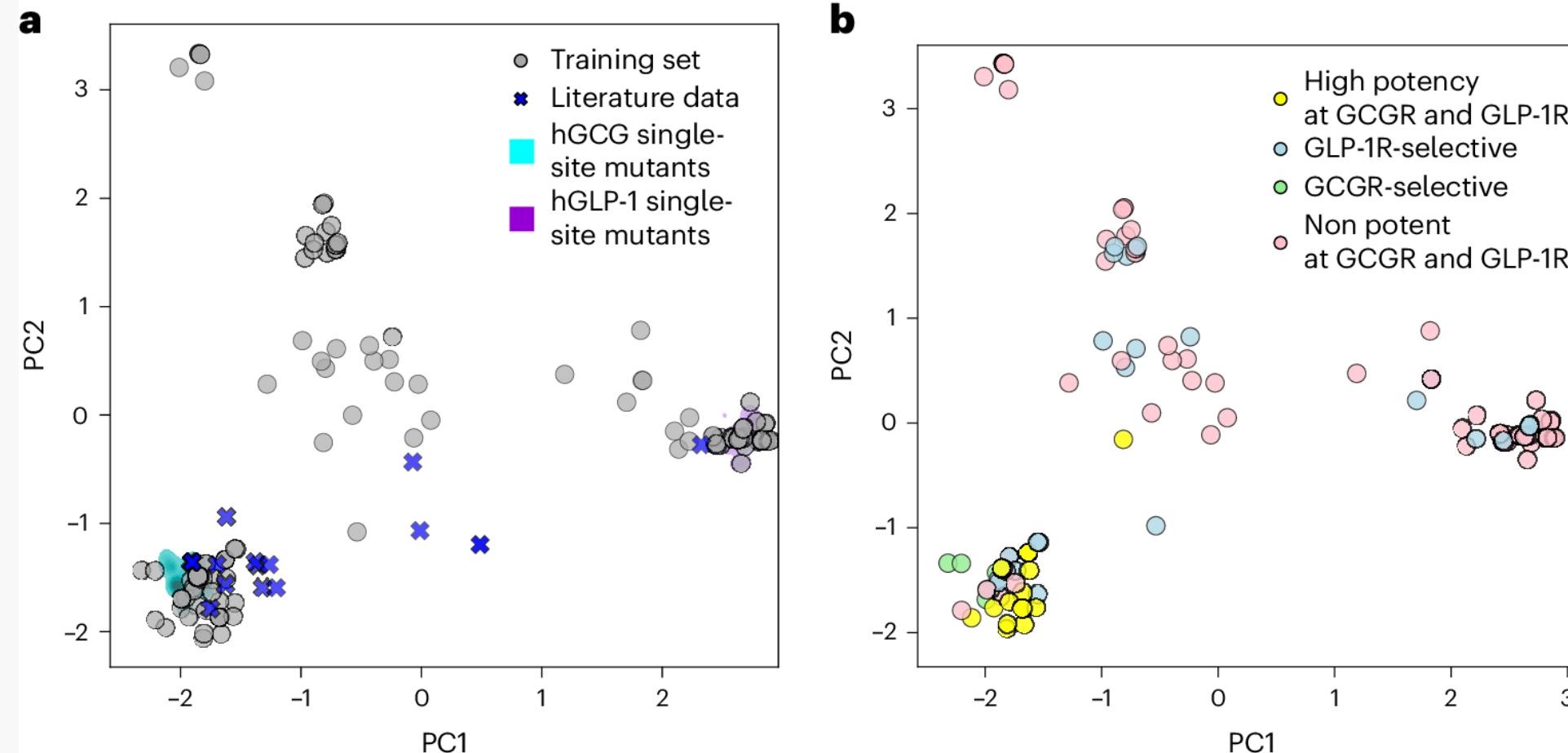
### グルカゴンの例

[H,S,Q,G,T,F,T,S,D,Y,S,K,Y,L,D,S,R,R,A,Q,D,F,V,Q,W,L,M,N,T,-]

ペプチド配列を学習可能な形式にエンコード



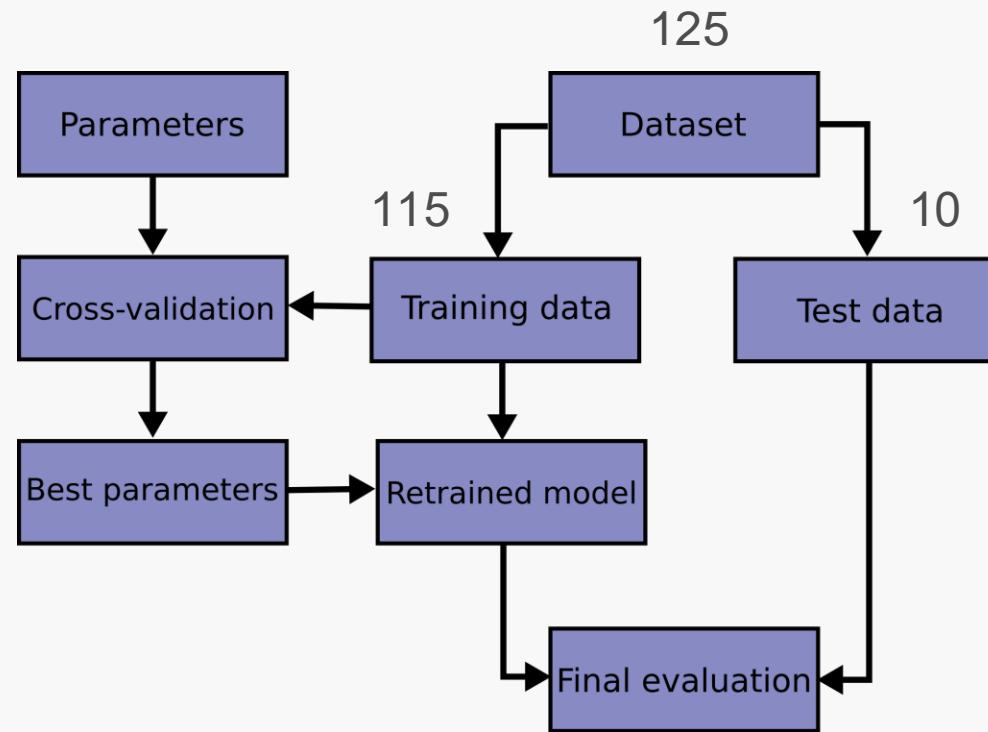
## 配列の類似性の観点でトレーニングデータの分布を示すために主成分分析



デュアルアゴニストにはhGCGとの配列類似性が必要な可能性

## 機械学習モデル構築・パラメータ調整の流れ

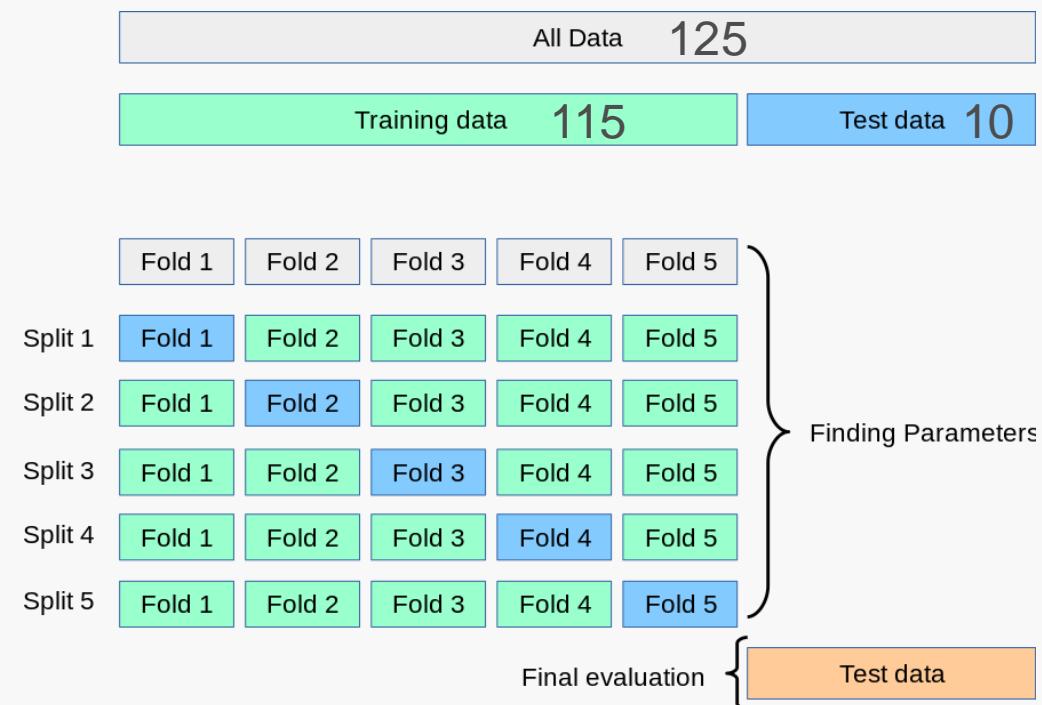
### モデル構築のフロー



[scikit learn 3.1.Cross-validation]

### 交差検証

ハイパー・パラメータ(人が設定しないといけない値)の調整



### 7つのモデルを構築し性能を比較

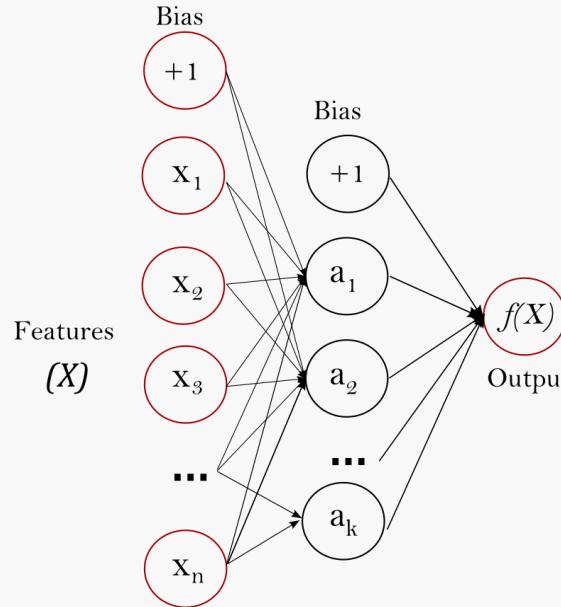
- 損失関数はr.m.s.eを用いてモデルの評価
- 外部データセットの測定値をテストデータとして用いたモデルの評価

$$r. m. s. e = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

Models	Cross-validation on training data		Validation on held-out literature data <sup>15</sup>	
	R.m.s.e. GCGR	R.m.s.e. GLP-1R	R.m.s.e. GCGR	R.m.s.e. GLP-1R
Ridge	0.63 ± 0.06	0.75 ± 0.10	1.12 ± NA	0.63 ± NA
SVR	0.60 ± 0.04	0.81 ± 0.07	1.09 ± NA	0.45 ± NA
GPR	0.69 ± 0.17	0.88 ± 0.23	1.25 ± NA	<b>0.25 ± NA</b>
Random forest	0.62 ± 0.04	0.77 ± 0.06	0.98 ± 0.05	0.61 ± 0.06
NN single-task	0.67 ± 0.07	0.78 ± 0.07	0.92 ± 0.18	0.41 ± 0.11
NN multi-task	0.68 ± 0.06	0.78 ± 0.05	1.04 ± 0.08	0.73 ± 0.36
NN multi-task ensemble	<b>0.59 ± 0.05</b>	<b>0.68 ± 0.04</b>	<b>0.86 ± 0.02</b>	0.53 ± 0.08

NN multi-task ensembleが一番予測精度が高いモデル

## NN(ニューラルネットワーク)

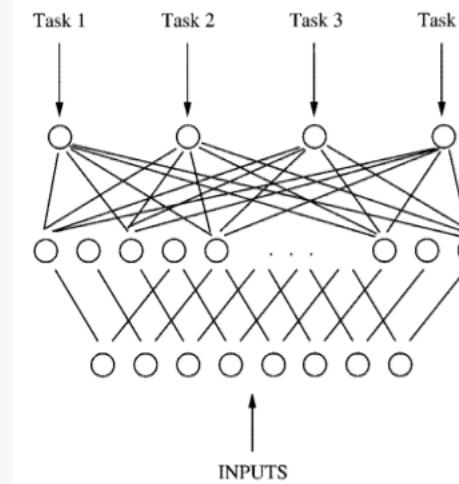


入力層 隠れ層 出力層

入力層から受け取った値に重みをつけた形で  
隠れ層のニューロンが受け取り、同様に隠れ層  
から重みをつけて次のニューロンへ値を送る  
方式で、最終的な予測を計算

[scikit learn 1.17. Neural network models]

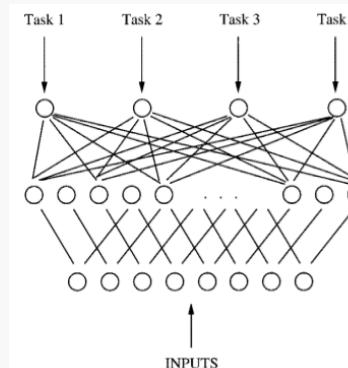
## マルチタスク学習



GCGR, GLP-1Rそれぞれに対して  
モデルを作るのではなく、一つのモデル  
で複数のタスクを解く手法  
1つのタスクに対して過度に適応する  
ことを防ぐことが可能

[Sebastian Thrun, et al. Machine Learning 1997]

## アンサンブル



複数のモデル出力の平均によって  
最終的な予測を決める手法  
 $\times 12$  NNは重みのランダムな初期化から  
モデル構築をするため、予測に少し違  
いが出る

Fig1. トレーニングデータの作成

Fig2. 機械学習モデルの構築と評価

Fig3. 作成したモデルを利用したリガンドのデザイン

Fig4. デザインしたペプチドの合成と評価

Fig5. トレーニングデータとデザインしたペプチドの比較

Fig3 予測するペプチドのデザインと活性予測

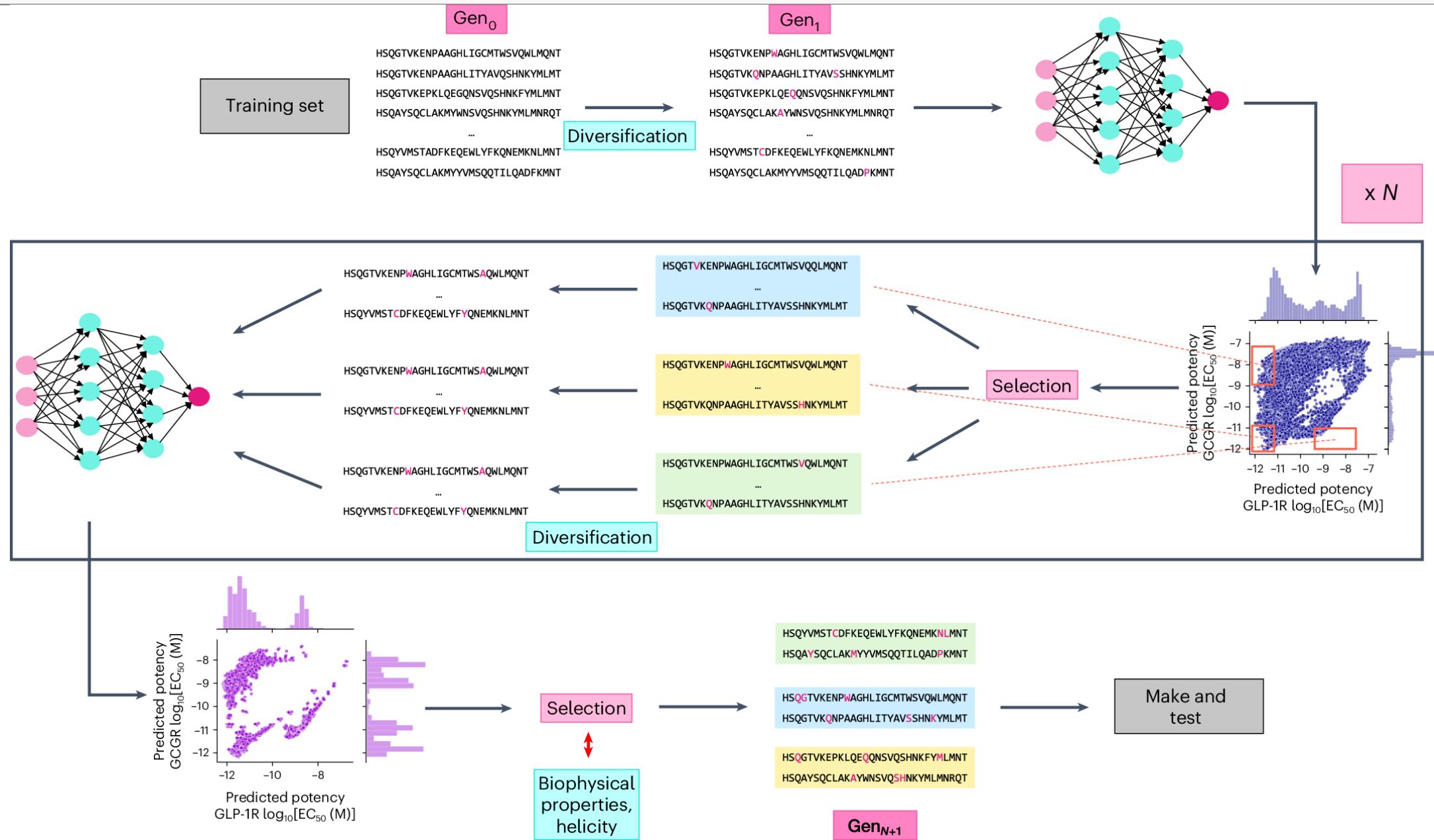


Fig3 予測するペプチドのデザインと活性予測

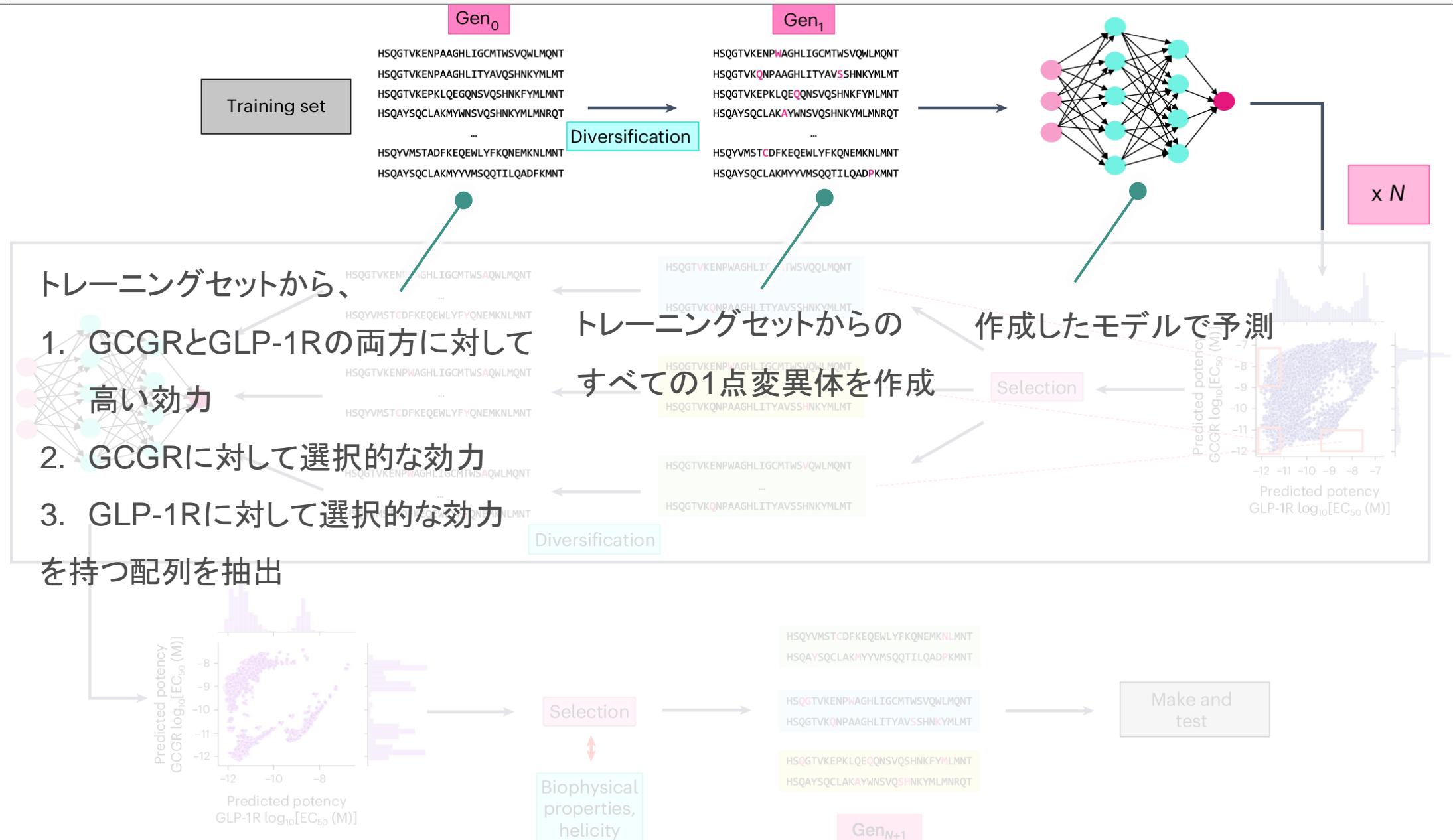


Fig3 Ligand design

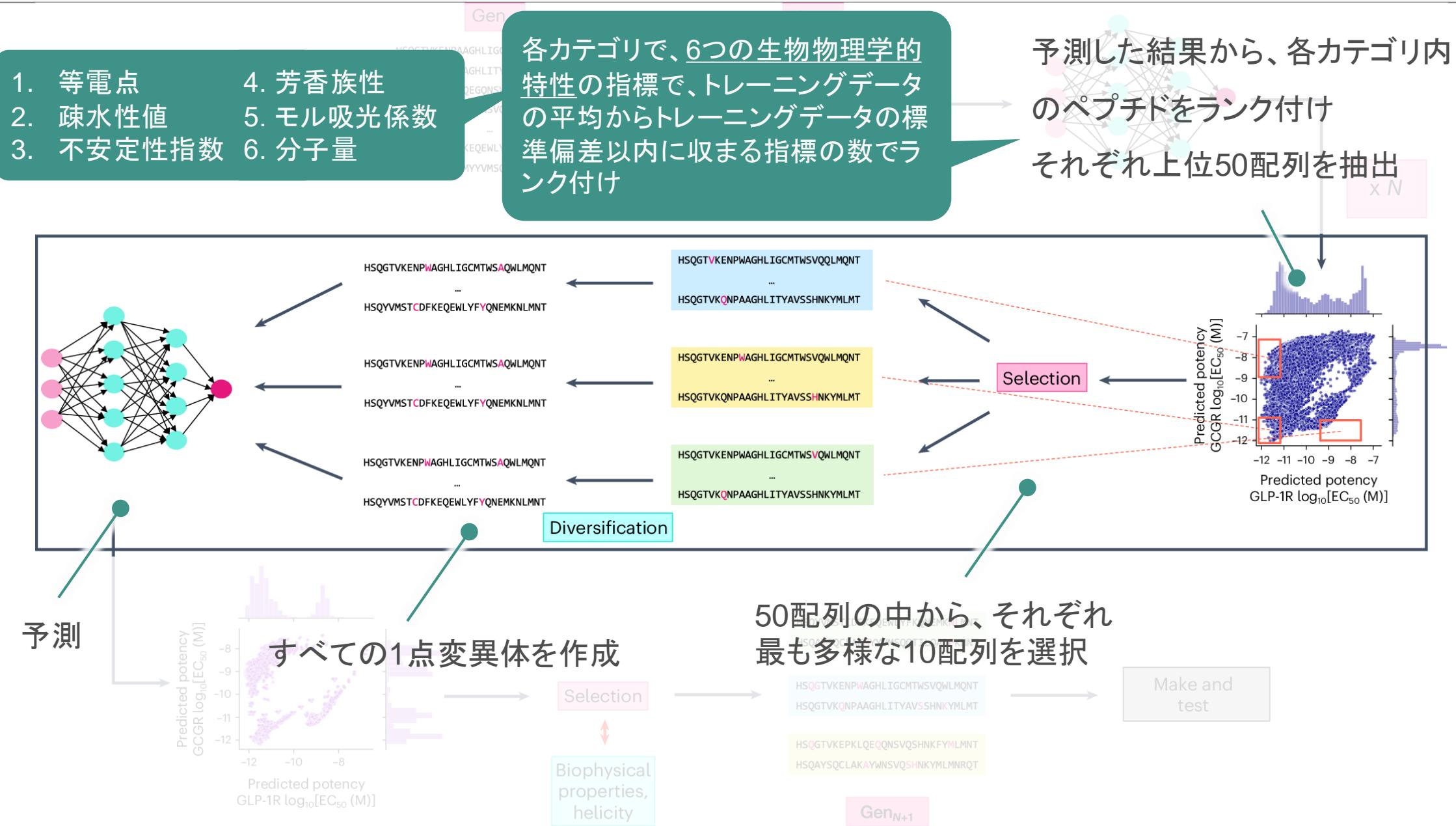
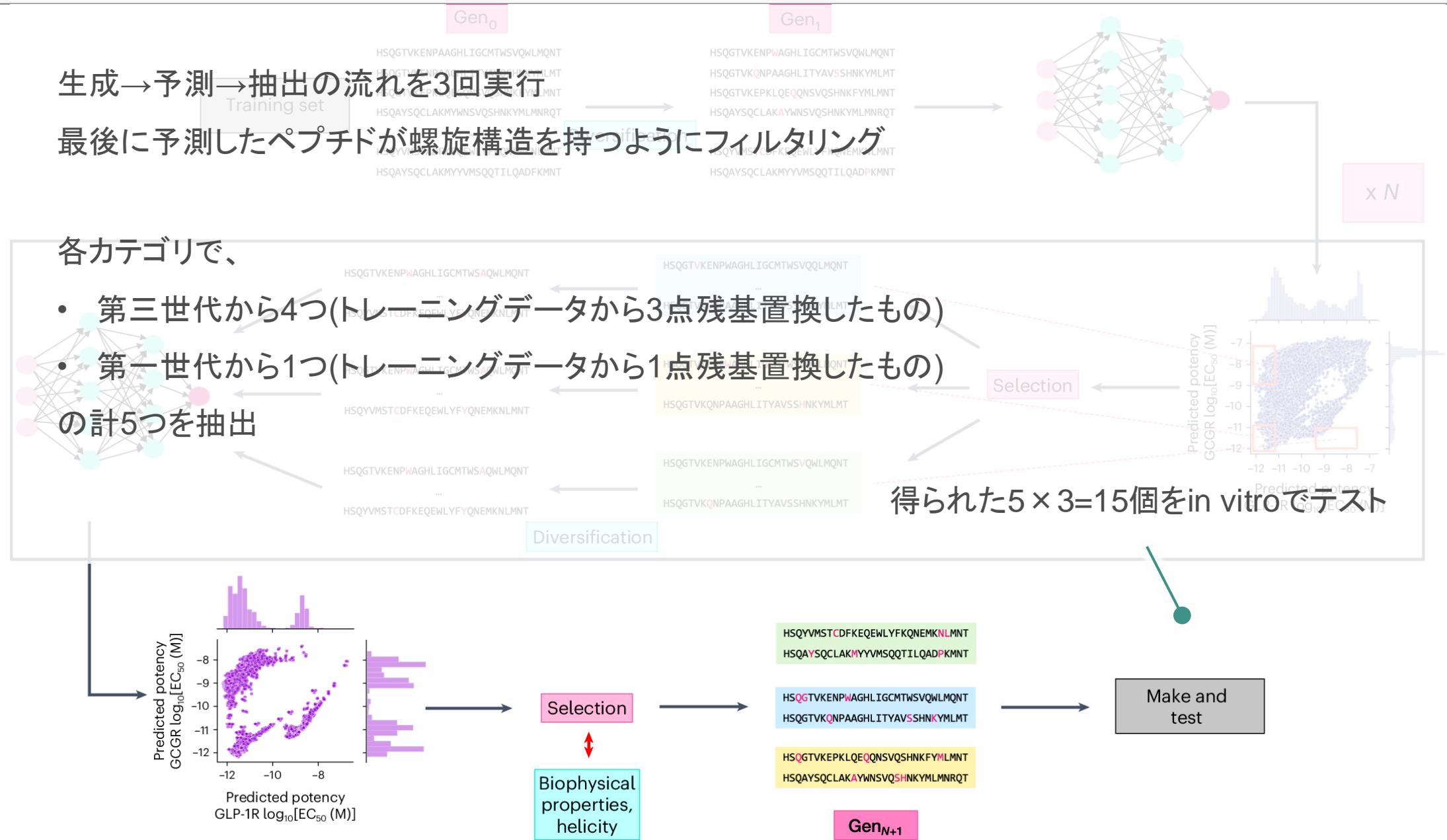


Fig3 Ligand design



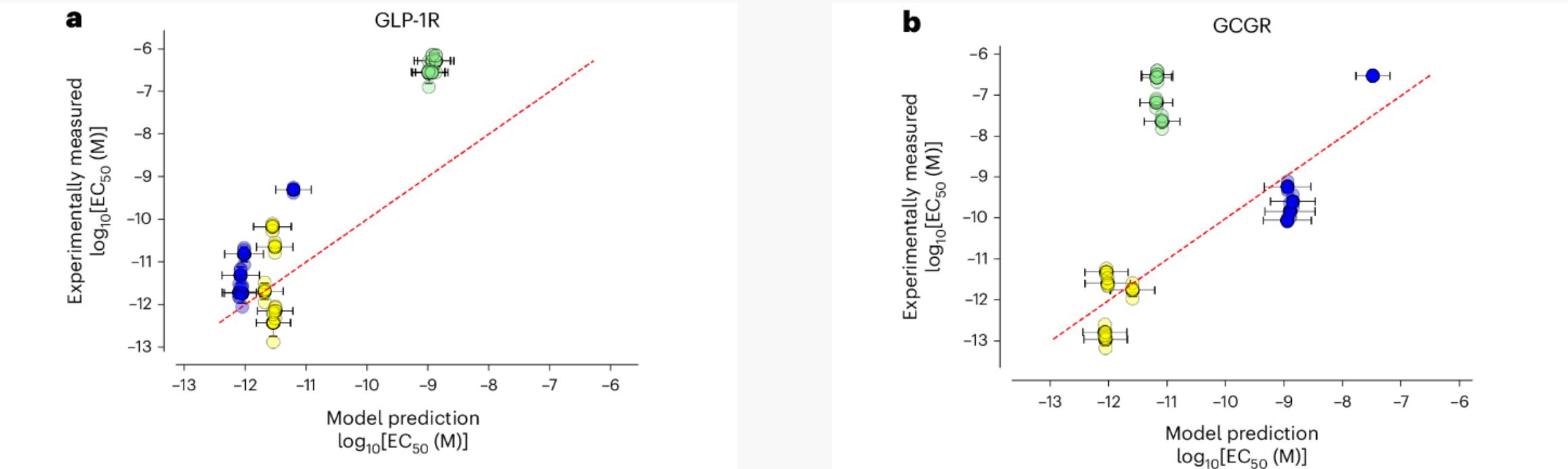
デザインした計15個のペプチドを合成して評価し予測値と比較

トレーニングデータ作成時と同様に、CHO細胞にhGCGR, hGLP-1Rを発現させ、  
ペプチドを添加してそれぞれのEC50値を測定

黄：デュアルアゴニスト

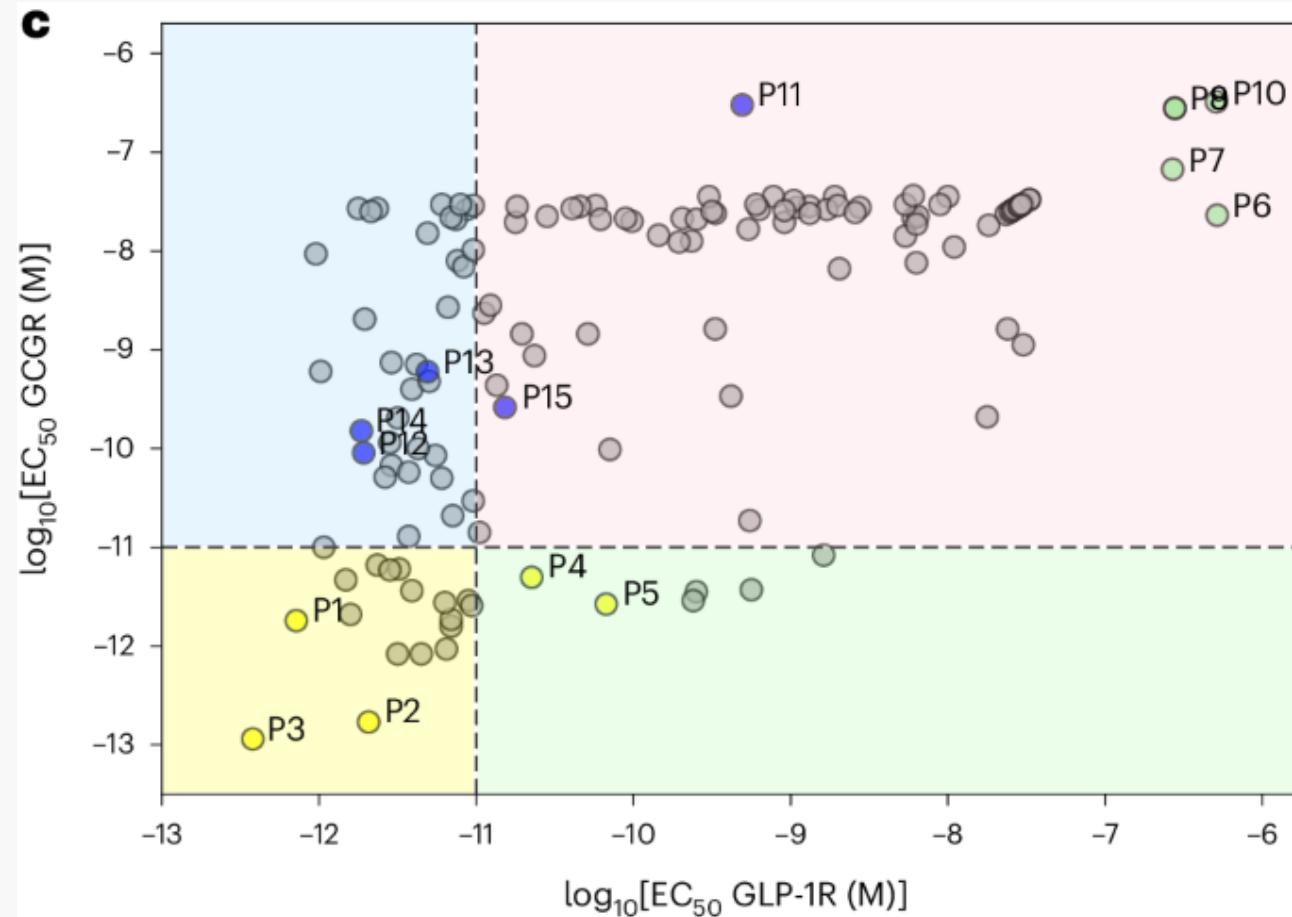
青：GLP-1R選択性

緑：GCGR選択性



デュアルアゴニストと、GLP-1R選択性的アゴニストで、正確な予測が確認

## デザインした計15個のペプチドとトレーニングデータの比較

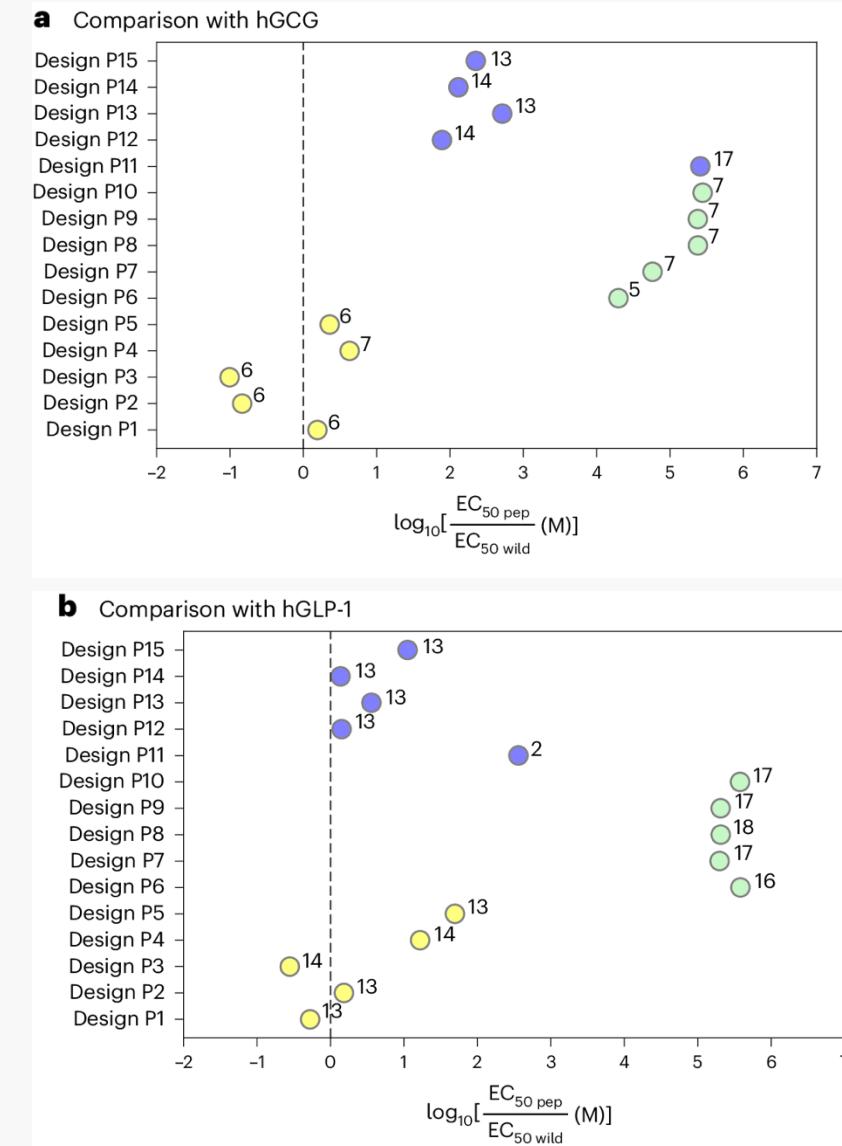


P1, P2, P3はどのトレーニングデータよりも両方の受容体で高い活性

Fig5 予測した配列の解析

- P1-P5
  - C末端領域に6-7の置換があるhGCGに近い類似体
- P6-P10
  - C末端領域に5-7個の置換があるhGCGに近い類似体
  - GCGR活性の喪失
    - 11位(S->D) 23位(V->G, Q, R) が全体的に見られた
    - P8の新しい変異: 20位:H, 22位:Q
- P1, P3
  - hGLP-1に比べ、P1は2倍、P3は3.6倍GLP-1R活性が高かった
    - GLP-1R活性化に重要な29位の変異(G->H)と(G->L)を持つ
- P11-P15
  - P11だけhGLP-1に近い変異で、GLP-1RのEC50を400倍近く低下させた
    - 11位(S->A), 24位(A->P)
  - P12-P15はhGLP-1に対して13残基の変更
    - GLP-1Rの活性化に重要な4,7,9,22,23位のアミノ酸を正しく保存

[Jung-Mo Ahn, et al. J. Med. Chem. 2015]



モデルは重要な配列と活性の関係を捉えていた

## 結言

機械学習によって、限られたデータにも関わらず、GCGRとGLP-1R両方に対する効力をデータセットの配列より同時に7倍以上向上させるデュアルアゴニストペプチドを発見した。

## 展望

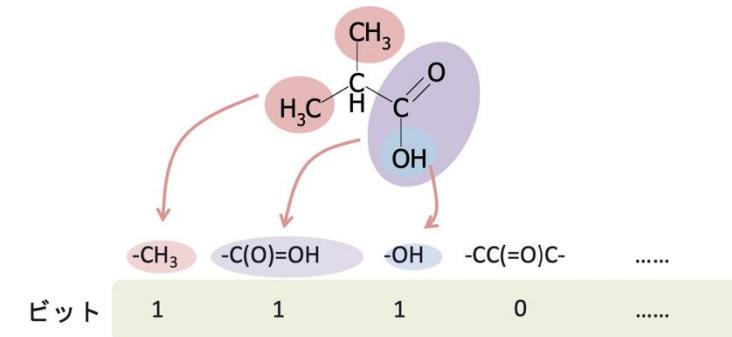
- より多くのデータを収集し、モデルの配列設計能力を向上させる
- GCGR/GLP-1R/GIPRトリプルアゴニストペプチドデザインへの応用

## 現在の研究 - 構造ベースのペプチド探索

分子フィンガープリントを用いて、部分構造の有無を入力データとした

TGR5活性予測

- 1配列あたりのデータ量(部分構造の数)が多く、過学習しやすい
- 予測根拠が分かりにくい



## ペプチド配列ベースの機械学習を用いたTGR5活性予測

ペプチドの配列と、TGR5のEC50のデータのセットを入力データとして機械学習モデルを構築する

### 利点

- 予測根拠が分かりやすい
- 構造ベースの過学習問題を解決できるかもしれない

### 懸念点

- 3残基と短い配列(小さいデータ)で十分な予測ができるか
- 多様なデータセットを得ることができるか

## 胆汁酸受容体のTGR5とFXR

核内受容体FXRは胆汁酸ホメオスタシスの主要な調節因子であるが、FXRは心筋虚血の際にアポトーシスを誘発して新筋細胞の生存率を低下させるという報告もある。

[Ming Xu et al. JMCC, 2021]

## マルチタスク学習によるTGR5選択的アゴニストペプチドの探索

TGR5をトランスフェクション



ペプチド添加



発光測定



ペプチド添加



FXR LBD, コアクチベーター  
ペプチド添加



TR-FRET測定



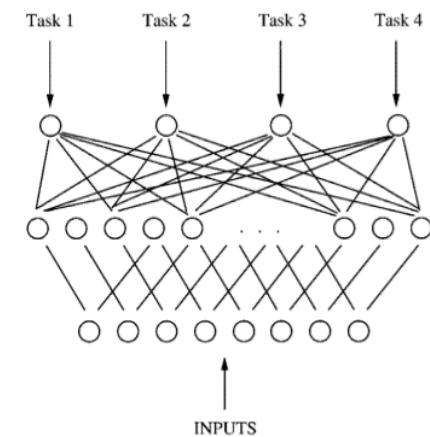
## TGR5選択的アゴニストの開発

- TGR5に結合して作動する
- FXRに結合せず活性を示さない

## マルチタスク学習による予測

TGR5活性のデータ

FXR活性のデータ



(TR-FRET FXR Coactivator Assay Kit)