日本知財学会　第17回年次学術研究発表会
2019年12月7日（大岡山）

パネル討論　～先端技術と知財戦略～

IT創薬と知財

秋山　泰
東京工業大学 情報理工学院

Tokyo Tech

秋山 泰／あきやま ゆたか

東京工業大学 情報理工学院　教授

東京工業大学 中分子IT創薬研究推進体（MIDL） 研究代表者

東京工業大学 リーダーシップ教育院　（ToTAL）　副教育院長

NPO法人 並列生物情報処理イニシアティブ (IPAB)　理事長

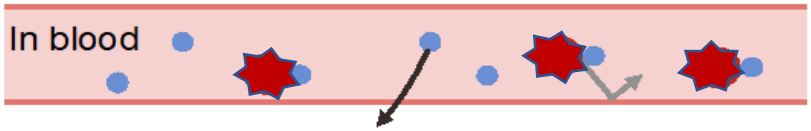
産業技術総合研究所　創薬分子プロファイリング研究センター　研究顧問

1990年　慶應義塾大学院工学研究科電気工学 博了 工学博士
1990年　通商産業省 工業技術院 電子技術総合研究所 研究官
1992年　京都大学 化学研究所 助教授
1996年　技術研究組合 新情報処理開発機構 並列応用研究室長
2001年　産業技術総合研究所　生命情報科学研究センター長
2007年　東京工業大学 大学院情報理工学研究科 教授

私が取り組んでいること

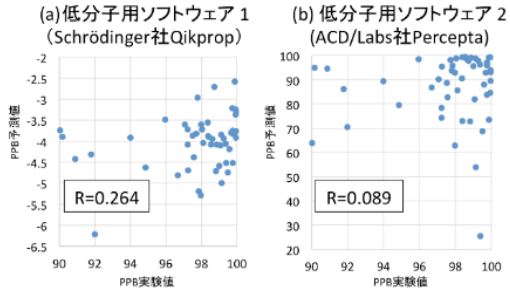
ペプチド医薬品のボトルネックを解消したい

問題点① 体内持続性が比較的低い



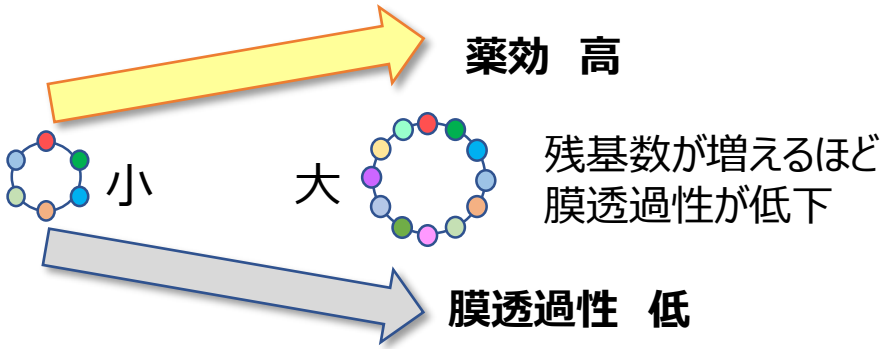
血中タンパク質結合率（PPB）が従来薬より低く、
短時間で体内から分解・排泄されてしまう

現状 既存技術では予測が不可能
試行錯誤による設計



体内持続性予測の事業化
（独自データ収集＋機械学習）

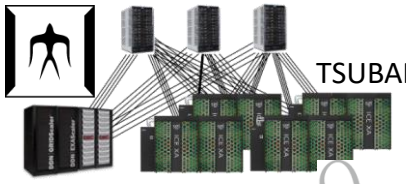
問題点② 細胞膜を通過しにくい

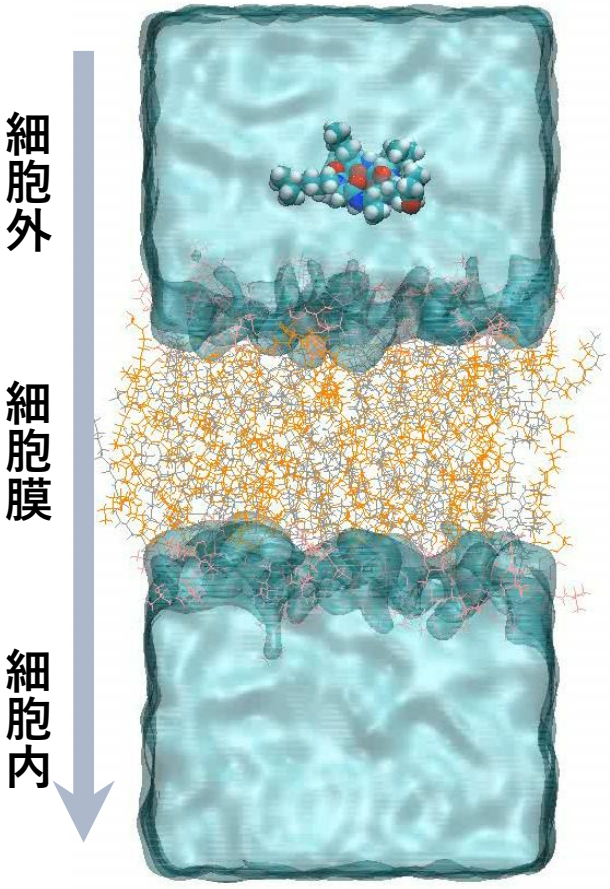


現状 既存技術では予測が不可能
試行錯誤では良い設計が見つからない

膜透過性予測の事業化
（独自データ収集＋機械学習＋
超高速分子シミュレーション）

機械学習と分子シミュレーション

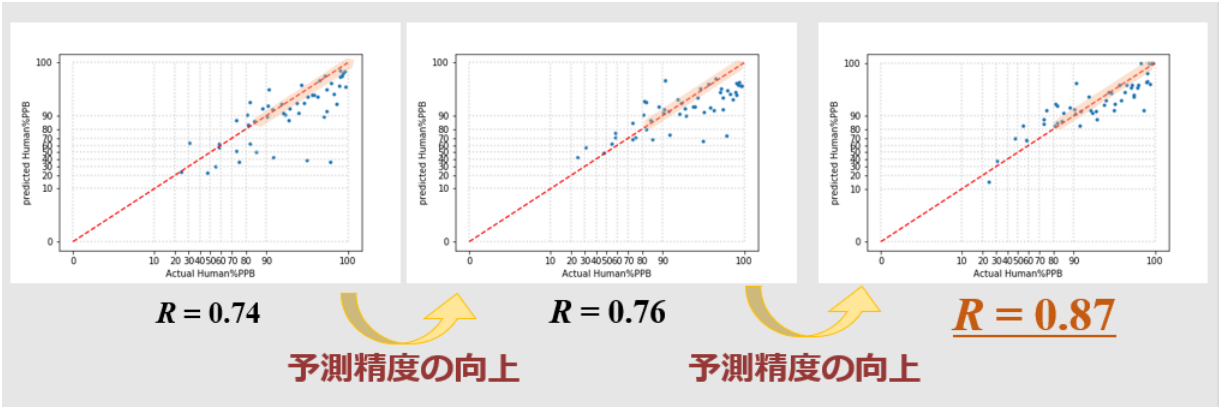




ペプチド膜透過シミュレーション
（GPUを最大800枚同時に使います）

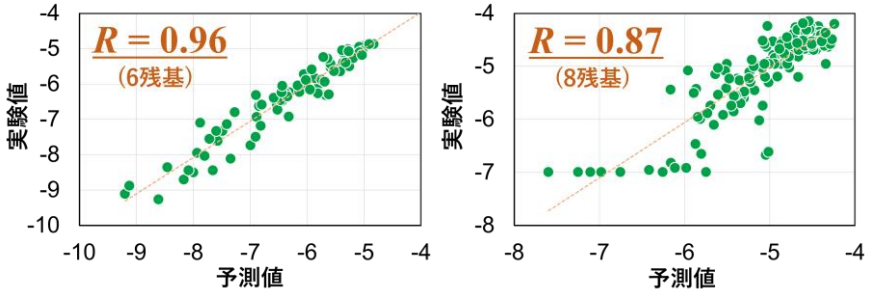
体内持続性予測システム

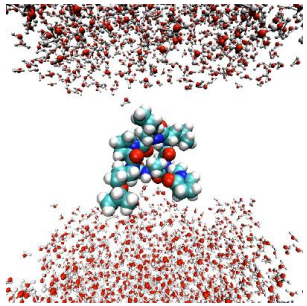
- 独自実験および企業連携で、**希少データ収集**
- **性質の全く異なる「低分子」データからも機械学習**
- 事業化すれば、顧客からも多くのデータが収集できる



膜透過性予測システム

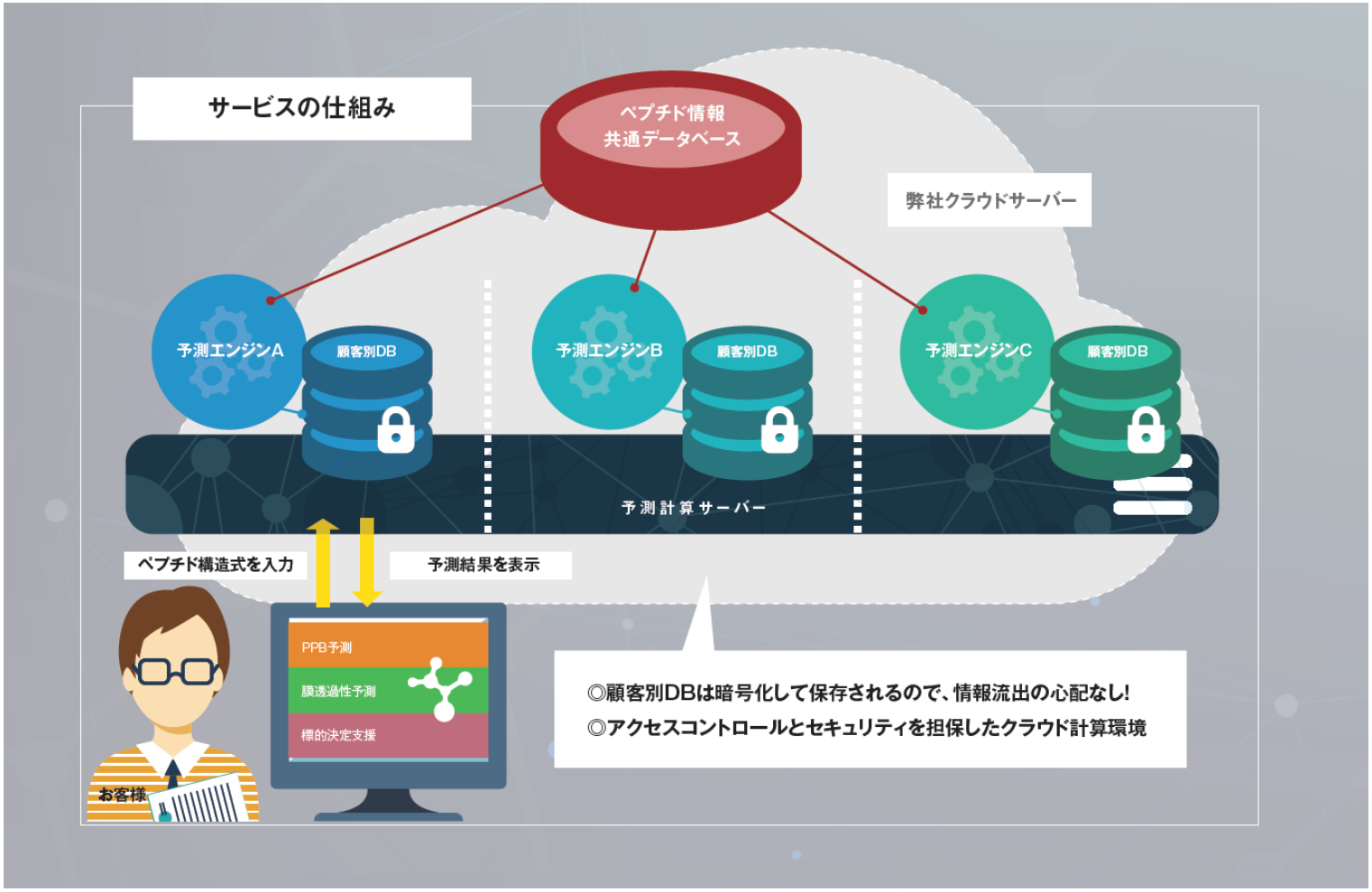
- **機械学習とシミュレーションの２つのアプローチを融合**
- 数ヶ月レベルの計算時間を、数時間へと超高速化





機械学習と相性が良いSaaS戦略

・学習モデルや教師データを日々アップデートできる（集中管理）
・契約に基づき、顧客の利用傾向を分析して改良を続けやすい
・ソフトウェアをリバース・エンジニアリングされにくい

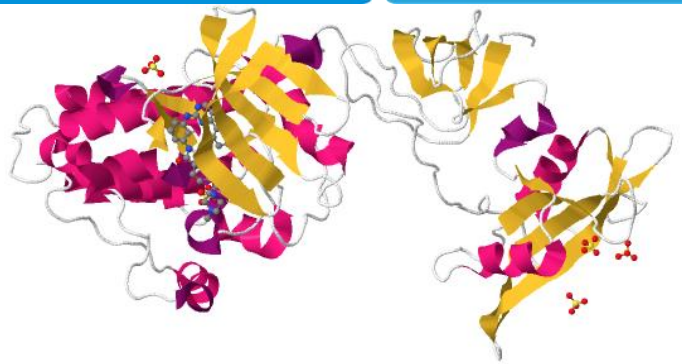


深層学習は創薬分野ではまだ苦戦している

（東工大　石田貴士准教授のスライドを一部改変）

- **IPAB　IT創薬コンテスト** (2014，2015, 2016, 2017)

  C-Yes酵素に阻害活性のある化合物を200万化合物ライブラリの中から予測。
  第1、2回でグランプリを受賞したグループはRandom Forestベースの手法。
  深層学習を用いたグループは低迷。



X-ray structure of c-Src (PDB ID: 1y57)

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Identification of potential inhibitors based on compound proposal contest: Tyrosine-protein kinase Yes as a target

Received: 07 August 2015
Accepted: 27 October 2015
Published: 26 November 2015

Enamine ID	Chemical Structure	Inhibition Rate %	Coefficient of variation %	Similarity	Filter class
Z820655914		89.0	22.0	0.75	LB&SB
Z126204226		37.2	35.1	0.77	SB
Z653349554		66.7	10.4	0.79	SB→LB

- Chiba et al., Identification of potential inhibitors based on compound proposal contest: Tyrosine-protein kinase Yes as a target, *Scientific Report*, 2015

深層学習を使うだけでは劇的な性能改善はない？

ビッグデータ　vs.　スモールデータ

ビッグデータ	スモールデータ
各データは**相似的なので貪欲に収集** （個人や都市の違いを越えて）	当該問題か類似例だけ**特化して収集** （異物を混ぜると性能が下がるよ）
１件あたりの収集コスト　＜**低**＞	１件あたりの収集コスト　＜**高**＞
予期せぬ応用が発見される可能性	集めたデータの**適用範囲は限定的**
オープン戦略で使っていくべき	**クローズド**で使われる傾向

機械学習は㊙レシピの塊

企業内
データ

記述子選択手法、
スパースモデリング

縮約化の有無

学習器の選択

実世界
ペプチド
膜透過性
とか

実験対象
実験手法
実験条件

独自
データ

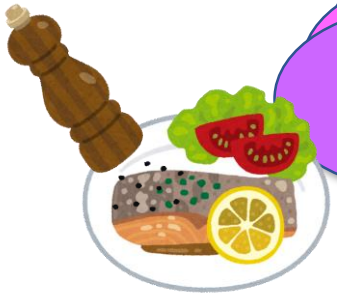
記述子
集合

Random Forest
Deep Learning
SVM, MLP, ...

入力表現の選択

文献・
特許情報

クレンジング
の方法の選択



結果の
解釈法

融合法の選択

アンサンブル学習
バギング、ブースティング

機械学習は㊙レシピの塊

特許で守れるの？

- システムを構成する各部品は汎用的な（数学的な）手法であって、「最終製品のための〇〇装置」のようなクレイムはしにくい。
  システムは必ずしも要素に還元しにくく、複雑な全体として捉えられる。

- **システムは常に変化**する。レシピ部分も変化するし入力データも変わる。
  どの段階で、何をもって特許を主張すればよいのだろうか。

- 手法やパラメータが漏洩すると、比較的**簡単にシステムを模倣**できる。
  模倣している事実について、外部からの**検証は著しく困難**である。
  形式的に異なって見えるが同等に機能するような修正も容易である。

→ **物質特許や製法特許とは異なる事情。**
ノウハウとして秘匿する方向に流れがち。

機械学習は㊙レシピの塊

ノウハウの秘匿で大丈夫？

- 提供する解答（例：創薬、材料組成、機械設計パラメタ）によっては**得られた１つの解の価値がシステム全体の価値より重い場合**がある。結果を活かして顧客側が個別に特許申請するまでは無保護なのか？

- 入力用の希少なスモールデータを秘匿して優位性を保つ戦略がある。しかし**データの漏洩に関する法的制度**は十分とは感じられない。

- **エンジニアの流動性**（リストラ…）の高さと、情報漏洩防止の相克

ペプチド創薬のための体内持続性や膜透過性の予測事業は
どうやって知的財産を保全すればよいのでしょうか？・・・