

「創薬とコンピュータ」

都市活力研究所/理化学研究所
(NPO法人バイオグリッドセンター関西)

志水隆一

2019年12月2日



1

1

目次

- 病気の原因
- 薬とは
- 遺伝子と蛋白質
- セントラルドグマ(遺伝子から蛋白質合成の流れ)
- 蛋白質の機能
- 蛋白質と薬
- 薬のストラテジー
- 副作用について
- 創薬におけるスパコンの利用
- バイオグリッドセンター関西の取組
- 実際の計算の概要
- 創薬と人工知能



2

2

病気の原因

- 蛋白質の暴走、怠慢
 - 蛋白質の活動が制御不能となる(アレルギー反応が止まらない、癌細胞の無限増殖)
 - 活動を抑制する薬: アントゴニスト
 - 蛋白質が機能しなくなる(血管新生が止まり細胞が壊死する)
 - 活動を促進する薬: アゴニスト



3

3

薬とは

一般的には疾患原因蛋白質を発見し、その蛋白質の機能を抑制する物質が薬となる。(再生医療においては、血管新生など機能を促進する場合もある)

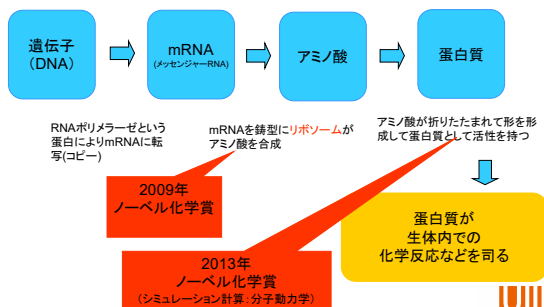
※成人病など原因が単一の蛋白質でなく、複合的な要因の病気も多数存在する。



4

4

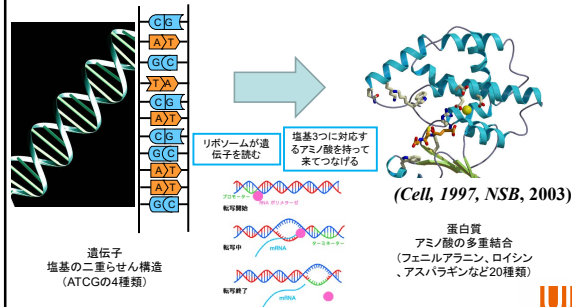
遺伝子から蛋白質の合成まで (セントラルドグマ)



5

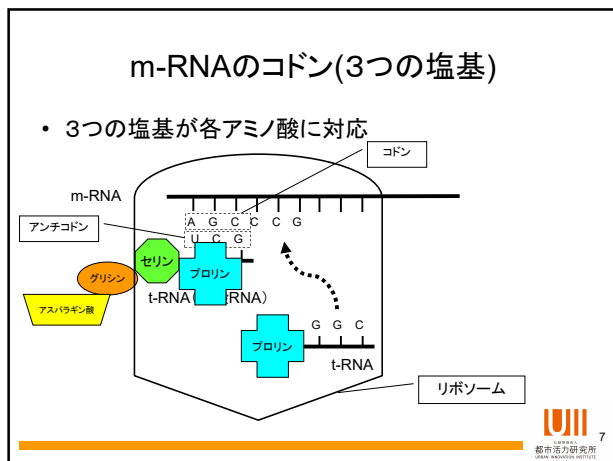
5

遺伝子から蛋白質 (セントラルドグマ)



6

6



7

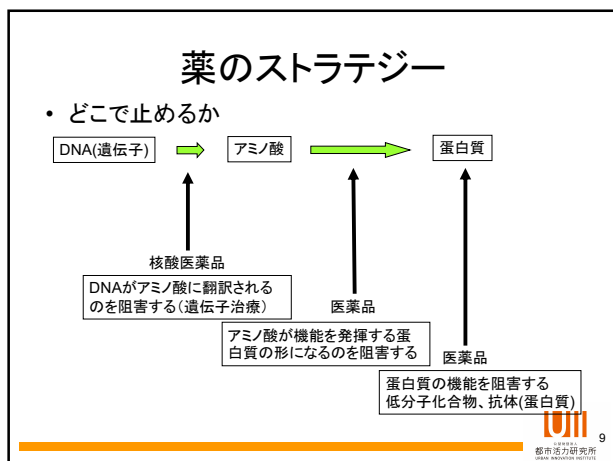
アミノ酸→蛋白質

- アミノ酸が多数連結するだけでは単なるアミノ酸の鎖(ポリペプチド)
- これがらせん状(α -ヘリックス構造)やジグザグ構造(β -シート構造)をとる。あるいは、2~4つの蛋白質同士が複合体を形成するなどし、蛋白質としての機能を発揮する。

ヘモグロビン: 4つのユニットが1つとなり酸素運搬の機能を担っている。

8

8



9

薬のストラテジー

-免疫チェックポイント阻害剤-

- 抗癌剤
- 新しい戦略(がんの免疫療法)
 - 免疫細胞ががん細胞を攻撃
 - がん細胞が免疫細胞のスイッチをOFF
 - 免疫細胞のスイッチを隠す(抗PD-1抗体)
 - PD-1 (programmed cell death-1)
- 「オプジーボ」(一般名ニボルマブ)
- 医療費: 年間3500万円⇒2018年から850万円に改定
- 開発: 小野薬品工業(2018年 ノーベル賞 本庶佑氏)

10

10

高額薬剤

- ギリアド・サイエンシズ社
 - 薬剤名: ハーボニー 対象疾患: C型肝炎 薬剤費: 500万円(3ヶ月)
- ノバルティス社
 - 薬剤名: キムリア 対象疾患: 抗がん剤 薬剤費: 3300万円(1回)
- ノバルティス社
 - 薬剤名: ラクスターナ 対象疾患: 網膜症 薬剤費: 9700万円(1回)
- ギリアド社(第一三共)
 - 薬剤名: イエスカルタ 対象疾患: リンパ腫 薬剤費: 4200万円(1回)

高額療養費制度: 月額負担額の上限が3.5万~25万円(収入による)
 例: 年収 370-770万 なら 8万円
 → 高額医療は保険の財政を圧迫する要因となっている。

11

11

DDS(Drug Delivery System)

目的の部位にどうやって薬を届けるか?

- 経口投与(飲み薬)
 - 主に腸で吸収して血管へ送られる
- 注射(患部注射、静脈注射)
 - 血流に乗って患部へ
- 塗り薬、貼り薬(皮膚浸透)
 - 直接患部へ

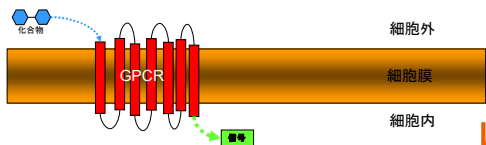
12

12

薬のストラテジーとDDS

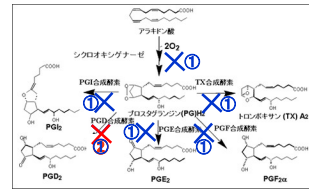
- ターゲット蛋白は分かった。では、どのように患部に届けるのか？
- 患部の臓器に届けるのも難しい。患部の臓器細胞の細胞内へとなるときにハードルが上がる。(だから核酸医薬は困難)
→ GPRCがターゲットとなる。

GPCR(G protein-coupled receptor):細胞膜を7回貫通している蛋白質。
5000個はあると推定。
2012年ノーベル化学賞 プライアン・コビルカ、ロバート・レフコウィッツ



13

副作用について



アレルギー情報伝達作用 胃粘膜粘液分泌促進作用 平滑筋(子宮・気管支・血管)収縮作用
胃の粘膜がなくなり胃が荒れる 腫瘍促進剤

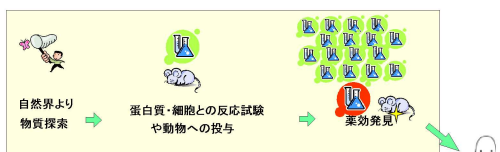
できるだけ直近②で止める方がいい。
①で止めるとPGE2もとまり「胃が荒れる」、PGF2αもとまる。

14

創薬手法の比較

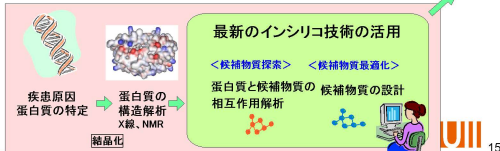
【従来法】(非効率な物質探索。化合物の大量消費→長期間および高コスト)

10-15年
500億



【新手法】(論理的かつ効率的)

期間短縮
開発費低減



15

これからの創薬

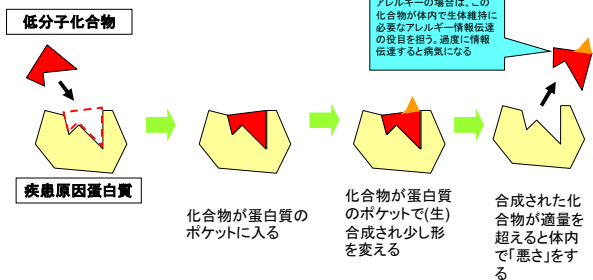
- 蛋白質の構造情報や蛋白質と化合物の反応情報などを統計的に解析し合理的・論理的な手法
- 情報の解析にコンピュータ(シリコンチップ)を使う



- 「インシリコ創薬」という
- (参考) 試験管: in vitro 生体: in vivo という

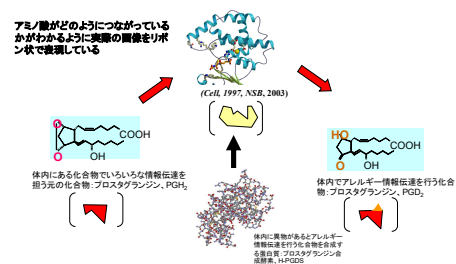
16

疾患のメカニズム



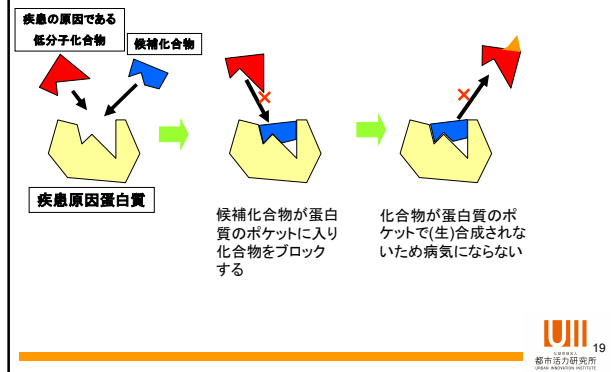
17

実際の画像



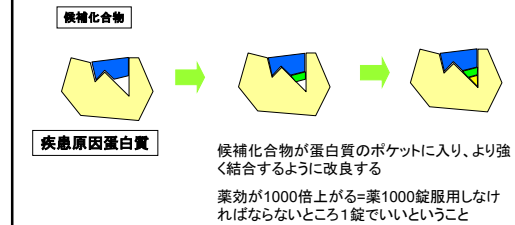
18

薬のメカニズム



19

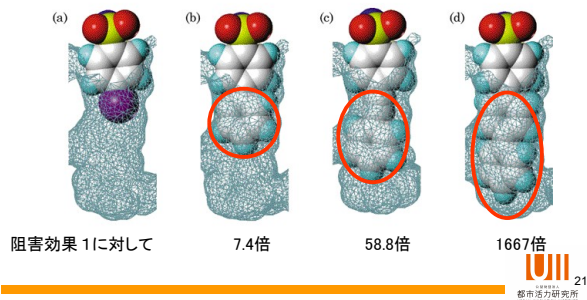
候補化合物の最適化



20

薬の構造の最適化

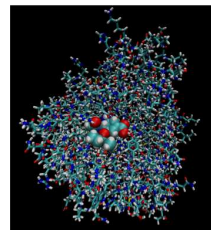
ポケットに対してピッタリ当てはまる大きさや形状を持った化合物が強い阻害活性を持つ。



21

創薬におけるスパコン利用の具体例

疾患原因蛋白質の活動を阻害する物質を蛋白質のシミュレーションにより探索する。
タミフルの開発はスパコンで行われた。



ウイルスが宿主細胞から別の細胞へと感染を広げる際に必要となるノイラミニダーゼというタンパク質を阻害する

インフルエンザの原因タンパク質に結合するタミフル

22

インシリコを実践しているコンソーシアム KBDD "K" supercomputer-Based Drug Discovery consortium



申請主体(事務局): NPO法人バイオグリッドセンター 関西
研究代表 京都大学大学院医学研究科 奥野昌史
製薬企業(19社):
アサヒカネ(1社): アサヒカネ(株)、旭化成ファーマ(株)、エーザイ(株)、森下博通(株)、(株)カネカ、キッセイ薬品工業(株)、各社製薬(株)、参天製薬(株)、塩野義製薬(株)、千寿製薬(株)、大正製薬(株)、田辺三菱製薬(株)、信人ファーマ(株)、草津(株)、日産化学(株)、日本新薬(株)、富士フイルム(株)、マルボ(株)、MeijiSeikaファルマ(株)
IT企業(2社): (株)東都コンテラ・テクノロジー・三井物産(株)
大学等: 京大医学部、徳島市大、兵庫医大、千葉大・医学部
(公財)都市活力研究所、産業技術総合研究所、理研、(公財)神戸医療産業都市推進機構

- コンピュータ創薬の根本課題に挑戦
- 製薬会社による現場利用に耐える計算フロー(計算精度と計算時間)の構築
- 我が国のコンピュータ創薬の中心拠点形成

23

プロジェクトの概要

- (1) 相互作用を予測する
蛋白質(キナーゼ、GPCR)600個と化合物3000万個の相互作用を計算
利用アプリ: CGBVS(京大・医・奥野)
Chemical Genomics Based Virtual Screening
- (2) ΔGを計算する
利用アプリ: MP-CAFE(東大・生産研・藤谷)
Massively Parallel Computation of Absolute binding Free Energy
- (3) Kon/Koffを計算する
利用アプリ: MSM(横浜市大)
Markov State Model

24

シミュレーション(第一原理計算)の限界

- 分子設計における相互作用
 - ファンデルワールス力
 - 静電相互作用(クーロン力)
 - 水素結合
 - これが全てではない!
- AIで解決
 - 野球のボールの軌道計算
 - 重力、空気抵抗
 - 厳密に計算するにはそのほかに風、雨、気温などを考慮する必要がある
 - AIによる予測が有効



25

25

人工知能とは

- 人工知能とは
 - 強いAIと弱いAI
 - 強いAI: AIが勝手に判断して答えを出す(鉄腕アトム)
 - 弱いAI: ある一定の学習の元で人間の精度を超える
 - ⇒ 現在のAIは弱いAI
- 弱いAIとはパターン認識技術である
 - パターンをコンピュータに学習させて新たな入力に対して結果を予測する



26

26

AIが脚光を浴びている理由

- 開発ツールが充実
 - Chainer Preferred Networks社(東大発ベンチャー)
 - Tensorflow Google社
 - Keras Tensorflowより簡単なツール
 - Arexnet 画像処理のツール
 - Python AIプログラミングでよく使われる言語(簡単)
 - GitHub 開発者の意見交換の場
- テストデータの充実
 - ImageNet 学習用: 120万画像、検証用: 5万画像、評価用: 10万画像
- コンピュータの発達
 - GPUマシン(Graphics Processing Unit) ゲーム用に開発されたCPU
 - » NVIDIA社



27

27

実際の開発

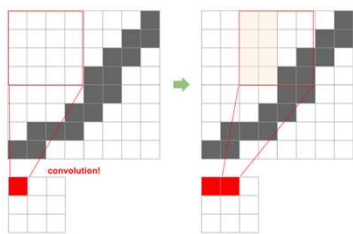
- 学習セットの準備
 - 入力と答えのデータの準備(教師あり学習)
 - ノイズなどの除去(これが大変)
 - ⇒ キューレーションとかクレンジングという
- プログラムのパラメータやネットワーク階層の設定
 - 重み付け、ネットワーク階層は人間が試行錯誤で決定(これがノウハウ)
- これら以外はツールを使えば数行のプログラミングでAIプログラムは完成する。



28

28

畳み込みのイメージ



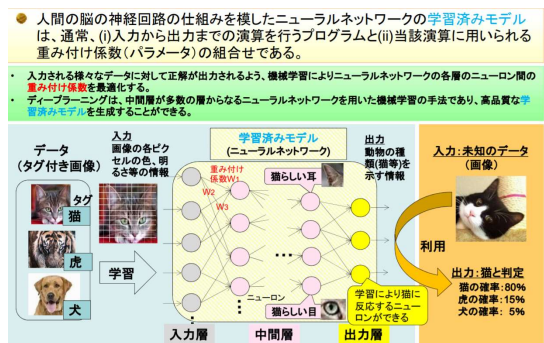
- 4×4のマスが特徴量により1つのマスに集約される
- 色をフィルタとするなら下の3×3の1番目は白、2番目はグレイと表現
- 次に、下の3×3を特徴量で集約し次の層を作成



29

29

AIの仕組み



出典: 特許庁調査基準室IoT関連技術の審査基準等について(平成29年3月)



30

30

何故、学習データと検証データを分けるのか？

- ・ 過学習をおこさないため
- ・ 過学習とは？
 - 学習データを過度に再現すること
 - データには、ノイズが含まれるがノイズまで学習してしまうことにより、学習データでは精度が高いが、予測したいデータでは、精度が落ちること

理化学研究所 LINC事務局 江口 浩平氏作成



31

31

人工知能の事例

– 具体例

- ・ グーグルの猫が有名
 - インターネット上の猫の画像を学習させ、新たに見せた画像が猫かそうでないかを判断させた
 - この学習にデープラーニング（正確にはCNN: Convolution Neural Network）という手法が使われた
- ・ 創薬ではLINC(Life Intelligence Consortium)
 - 蛋白質と化合物の結合ボーズを学習
 - » 新たな蛋白質と化合物の結合ボーズを予測
 - 腎臓における病理画像で壊死した部分を学習
 - » 新たな病理画像における壊死部分を予測した。
 - 再生医療分野で文献情報で有力な研究者を学習
 - » 有望な研究者を予測した。



32

32

シームレスなAI創薬プラットフォームをめざして



33

33

LINCのWGおよびPJの一覧

WG1. 未病・先創医療	WG5. モデル・分子設計・ADMET
1. 健康診断データによる発症予測	15. 合成経路予測
2. マイクロバイオーム・オミクスデータ解析	16. 分子設計AI
3. デジタルヘルス	17. 化合物記述子表現
	18. QSAR/QSPR/ in vitro ADMET予測
WG2. 臨床・診断	WG6. トランスレーショナルリサーチ
4. ゲノム医療におけるAI活用	19. 非臨床データからのヒトADMET予測
5. シミュレーションによる細胞分離	20. 疾患メカニズム解明・ブリッジング予測
6. AIによる病理画像処理	
7. AIによる電子カルデ処理	WG7. バイオリジクス・製剤・ロボティクス
WG3. 創薬テーマ創出	21. バイオリジクス関連AI
8. 有望提案先や研究テーマの自動探索	22. 結晶形・製剤関連AI
9. 標的分子探索	23. 調剤ロボティクス
10. ドラッグリポジショニング	WG8. 治験・市販後・メディカルアフェアーズ
WG4. 分子シミュレーション	24. AIによる治験の効率化
11. タンパク質立体構造・機能予測	25. 有害事象の情報基盤
12. AIによるドッキング計算高度化	26. 製剤Q&Aシステム
13. 分子動力学計算におけるAI活用	27. アウトカムリサーチ・医療技術評価
14. AIを用いた高精度分子カ場	WG9. 知識ベース・NLP (PJ28)
	WG10. AI基盤 (PJ29)



34

34

人工知能の課題

- これまでの科学の発展の段階
 - ・ 理論科学、実験科学、計算科学（コンピュータによる）
 - ・ 第4のアプローチとしてビッグデータ科学
- ビッグデータ科学
 - ・ 原因と結果の関係が不明
 - なぜかわからないが答えは合っている。
 - » この疾患にはこの化合物が効くが、なぜかは不明
 - » 製薬業としてはOKであるが、気持ち悪い
 - さらに教師なし学習法では
 - » コンピュータが自動でクラスタリングし、似たデータを抽出する。⇒人が見ても意味わからない。その結果「赤いスポーツカーに乗っている人は風邪を引きにくい」といった答えが予想される。
 - ・ 予想される結果から因果関係の仮説を立てる科学へ



35

35

ライフで利用するうえでのAIの課題

- データの収集
 - ・ 多種多様なライフ系のデータがあるがどのように収集するか？
 - ・ データ数、データの精度、データ項目（標準化）
 - 実験で入手困難な場合はシミュレーションデータで学習
 - AIによる辞書でのノイズの除去
- データの表現方法（計算機で認識できる表現）
 - ・ 「画像はセルをRGBの数値で表現できる」
 - ・ 化合物、蛋白質をどのように表現するか？
 - ディスクリプター、フィンガープリント
 - ・ 疾患名「肺がん」を数値ではどう表現する？
 - ・ その他 画像以外のデータの数値化



36

36