

「創薬とスパコン」

NPO法人バイオグリッドセンター関西
(都市活力研究所/理化学研究所)

志水隆一

2017年12月11日



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



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: アンタゴニスト
 - 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

薬とは

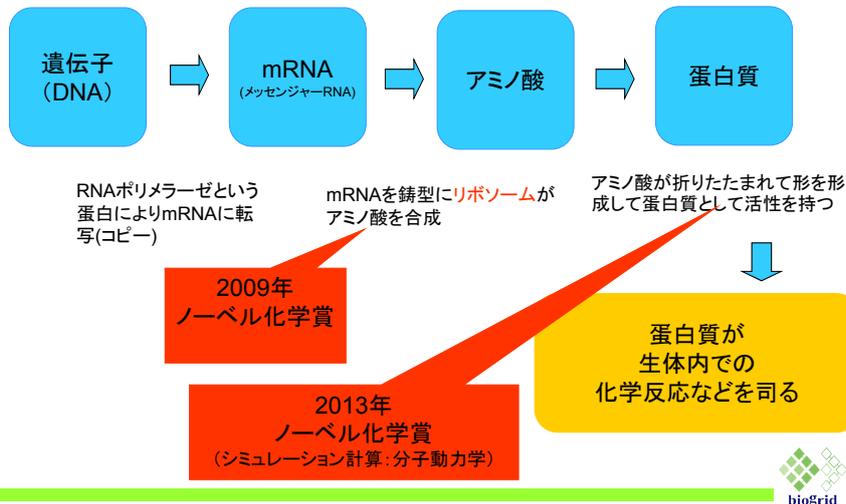
一般的には疾患原因蛋白質を発見し、その蛋白質の機能を抑制する物質が薬となる。(再生医療においては、血管新生など機能を促進する場合もある)

※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。

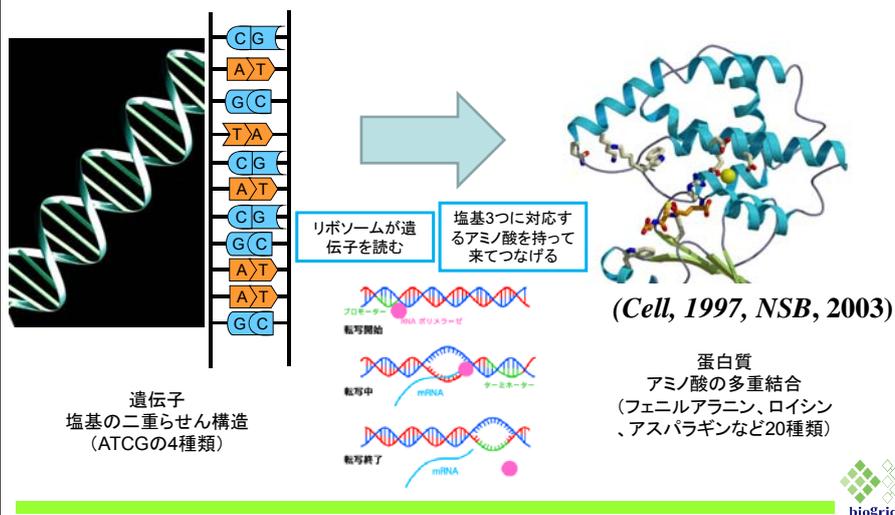


4

遺伝子から蛋白質の合成まで (セントラルドグマ)



遺伝子から蛋白質 (セントラルドグマ)



遺伝子と蛋白質

- 遺伝子は4つの塩基で構成(ATCG)
 - A:アデニン、G:グアニン、T:チミン、C:シトシン(U:ウラシル)
 - 螺旋状の2本の鎖で構成
 - A-T C-G が相互に結合
- 3つの塩基が20種のアミノ酸に翻訳
 - AAA→リジン、GAG→グルタミン酸、GAC→アスパラギン酸
 - GCC→アラニン、AUG→メチオニン など
- アミノ酸が連結して蛋白質を構成

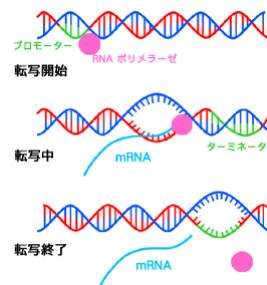
※4文字では表現できることに限界があるが、20種類では多様な表現が可能(アルファベットは26文字で、小説、論文、手紙など多彩な表現を実現)



7

遺伝子(DNA)→mRNA

- 遺伝子 4種類の塩基(A:アデニン、T:チミン(U:ウラシル)、G:グアニン、C:シトシン)より構成されており、2重のらせん構造をしている。
- AとT(U)、GとCが対になり結合
- RNAポリメラーゼが2重らせんをほどく
- DNAの塩基に対になるようにmRNAを合成 (AとU(ウラシル、DNAの場合はT)、GとC)



大学共同利用機関法人webより転載



8

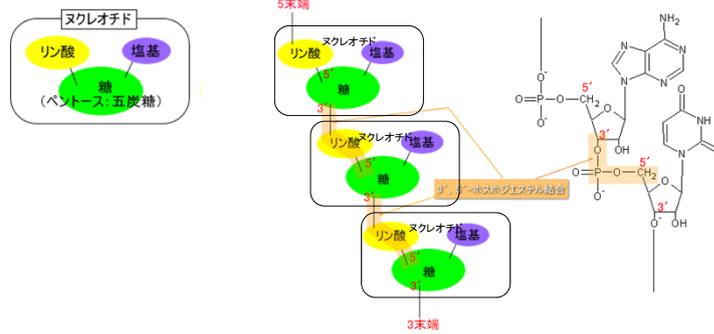
4種類の塩基 詳細(1)

ヌクレオチド(4種類の塩基+ペントース(ペン=5の意味)+リン酸)

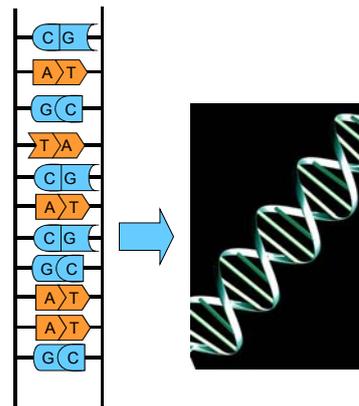
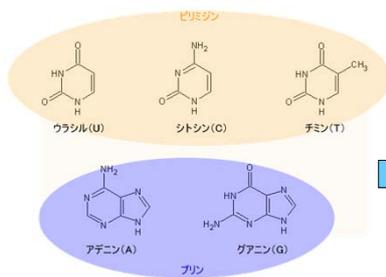


DNA, RNA(ヌクレオチドが重合)

塩基 これが4種類ある

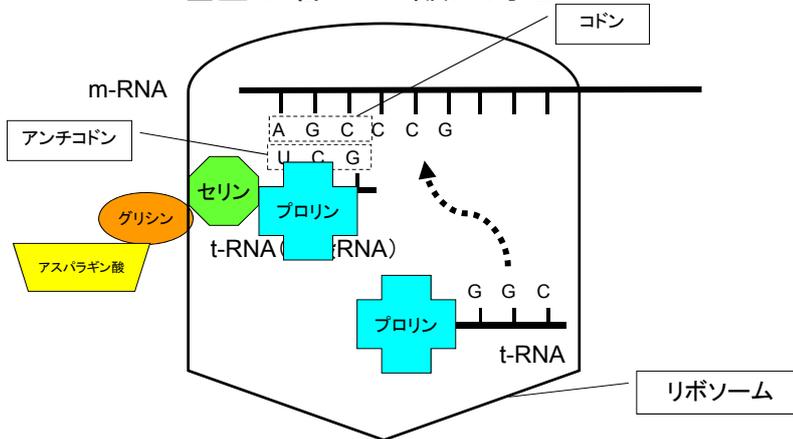


4種類の塩基 詳細(2)



m-RNAのコドン(3つの塩基)

- 3つの塩基が各アミノ酸に対応



塩基配列とアミノ酸(参考)

UUU (Phe/F)[フェニルアラニン](#)
UUC (Phe/F)[フェニルアラニン](#)
UUA (Leu/L)[ロイシン](#)
UUG (Leu/L)[ロイシン](#)

CCU (Pro/P)[プロリン](#)
CCC (Pro/P)[プロリン](#)
CCA (Pro/P)[プロリン](#)
CCG (Pro/P)[プロリン](#)

AAU (Asn/N)[アスパラギン](#)
AAC (Asn/N)[アスパラギン](#)
AAA (Lys/K)[リシン](#)
AAG (Lys/K)[リシン](#)

mRNA→アミノ酸(参考)

- リボソームがmRNAを読み込み、順番に3つの塩基に対応したアミノ酸をつなげる。塩基3つが1組で1つのアミノ酸を指定している。以下が3つの塩基と20のアミノ酸の対応表で「コドン表」と呼ばれている。

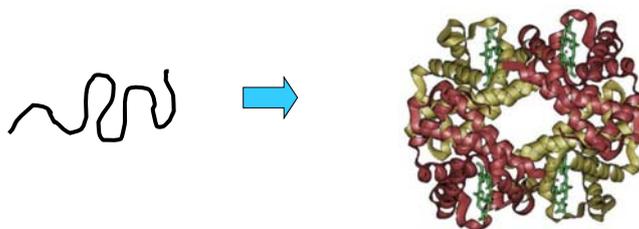
		2文字目					
		U	C	A	G		
1文字目	U	フェニルアラニン	セリン	チロシン	システイン	3文字目	
	U	フェニルアラニン	セリン	チロシン	システイン		U
	U	ロイシン	セリン	終止	終止		A
	U	ロイシン	セリン	終止	トリプトファン		G
	C	ロイシン	プロリン	ヒオスチジン	アルギニン		U
	C	ロイシン	プロリン	ヒオスチジン	アルギニン		C
	C	ロイシン	プロリン	グルタミン	アルギニン		A
	C	ロイシン	プロリン	グルタミン	アルギニン		G
	A	イソロイシン	トレオニン	アスパラギン	セリン		U
	A	イソロイシン	トレオニン	アスパラギン	セリン		C
	A	イソロイシン	トレオニン	リシン	アルギニン		A
	A	メチオニン	トレオニン	リシン	アルギニン		G
	G	バリン	アラニン	アスパラギン酸	グリシン		U
	G	バリン	アラニン	アスパラギン酸	グリシン		C
	G	バリン	アラニン	グルタミン酸	グリシン		A
	G	バリン	アラニン	グルタミン酸	グリシン		G



13

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)
- これらがらせん状(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。



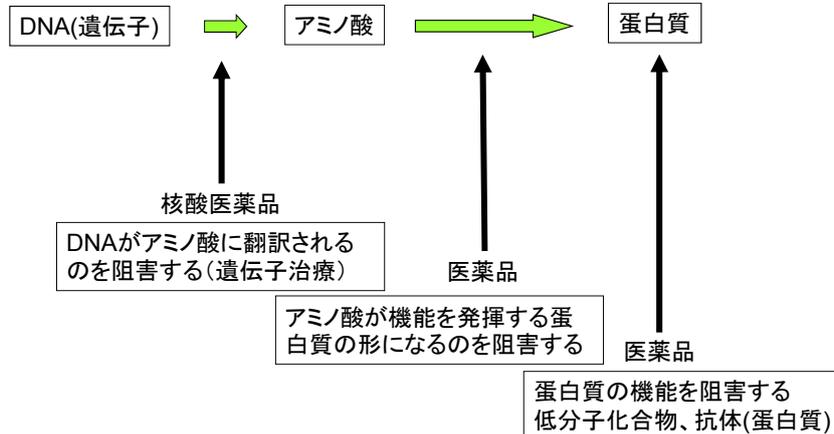
ヘモグロビン:4つのユニットが1つとなり酸素運搬の機能を担っている。



14

薬のストラテジー

- どこで止めるか



15

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⇒2017年から1700万円に改定
- 開発: 小野薬品工業



16

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:700万円(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:5200万円(1回)



17

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

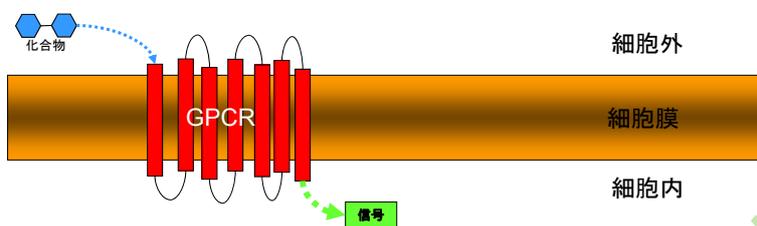


18

薬のストラテジーとDDS

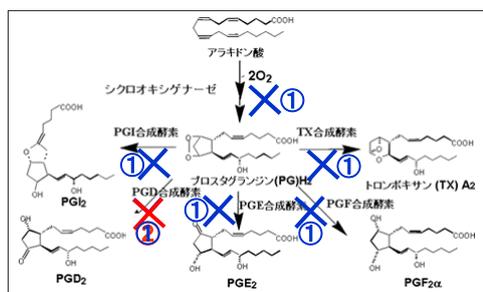
- ターゲット蛋白は分かった。では、どのように患部に届けるのか？
- 患部の臓器に届けるのも難しい。患部の臓器細胞の細胞内へとなるとさらにハードルが上がる。(だから核酸医薬は困難)
→ GPCRがターゲットとなる。

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



19

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 陣痛促進剤

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。



20

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

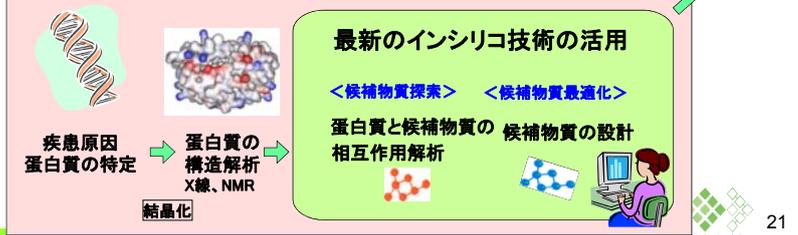
500億



【新手法】(論理的かつ効率的)

期間短縮

開発費
低減



bioGrid 21

創薬の工程

- これまで(実験中心の創薬)
 - 自然界(アマゾン、深海、東京湾)より物質を探索
 - 蛋白質と反応させ偶然ヒット
 - 化合物を最適化
 - 臨床開発
- 今後(コンピュータを使った創薬=インシリコ創薬)
 - 蛋白質の構造解析
 - コンピュータで設計
 - 化合物を合成
 - 化合物を最適化
 - 臨床開発

この過程がばくち的
20000分の1の確率

欧米ビッグファーマは物量作戦で
じゅうたん爆撃的に探索
化合物数
欧米製薬企業 50-100万
日本製薬企業 20-50万

従来手法に比べると非常
に合理的・論理的

bioGrid 22

スパコンを利用した創薬の工程

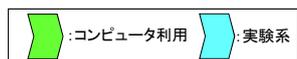
構造ベースのアプローチ



構造未知のアプローチ



何回か繰り返す



23

創薬におけるスパコンの利用の現状

- 化合物の探索
 - 蛋白質の構造に基づいた化合物の探索
 - SBDD(Structure Based Drug Discovery)
 - 実験データに基づいた化合物の探索
 - 機械学習法など
- 化合物の最適化
 - 基本骨格を決定した後、化合物の薬効向上
 - 物性改善(水溶性向上など)



24

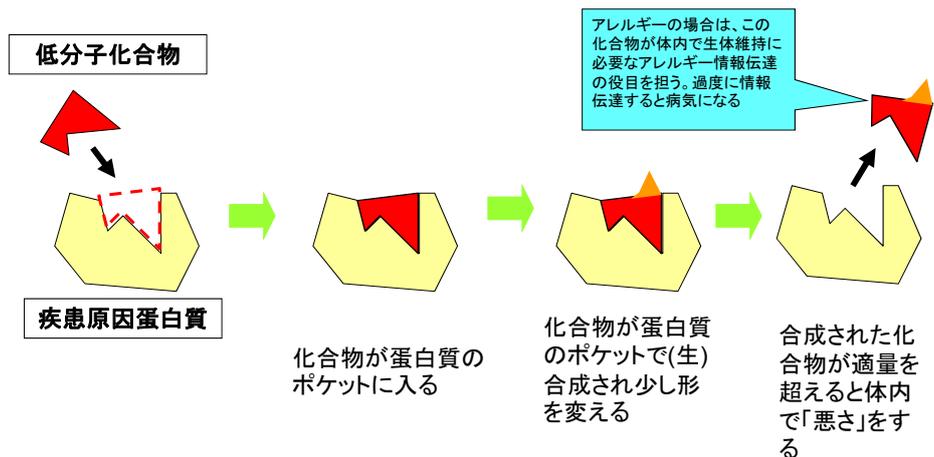
これからの創薬

- 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



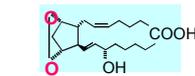
- 「インシリコ創薬」という
- (参考) 試験管:in vitro 生体:in vivo という

疾患のメカニズム

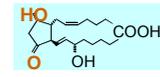
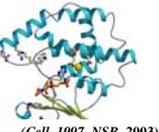


実際の画像

アミノ酸がどのようにつながっているかわかるように実際の画像をリボン状で表現している



体内に異物があるとアレルギー情報伝達を行う化合物を合成する
 蛋白質: プロスタグランジン合成酵素, H-PGDS

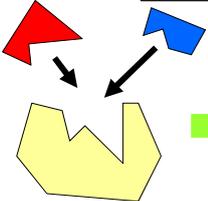


体内に異物があるとアレルギー情報伝達を行う化合物を合成する
 蛋白質: プロスタグランジン合成酵素, H-PGDS

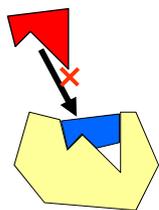
薬のメカニズム

疾患の原因である
 低分子化合物

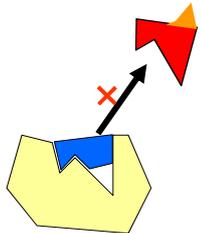
候補化合物



疾患原因蛋白質



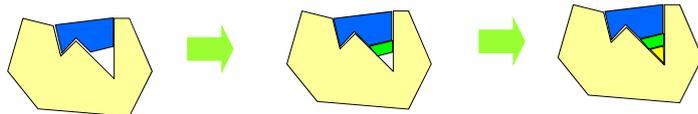
候補化合物が蛋白質のポケットに入り化合物をブロックする



化合物が蛋白質のポケットで(生)合成されないため病気になる

候補化合物の最適化

候補化合物



疾患原因蛋白質

候補化合物が蛋白質のポケットに入り、より強く結合するように改良する

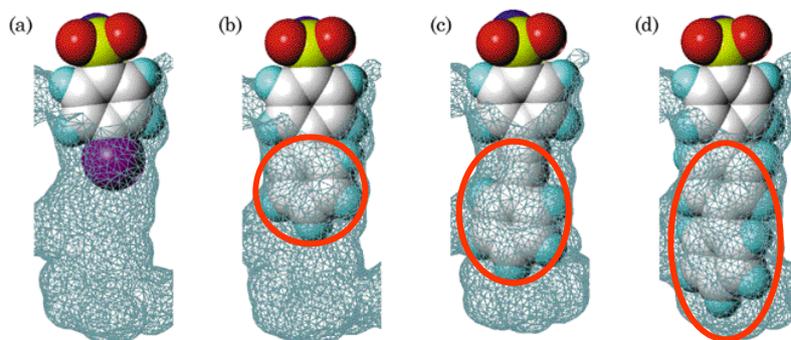
薬効が1000倍上がる=薬1000錠服用しななければならないところ1錠でいいということ



29

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。



阻害効果 1に対して

7.4倍

58.8倍

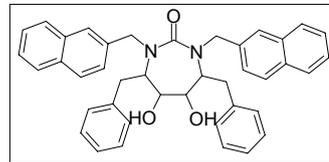
1667倍



30

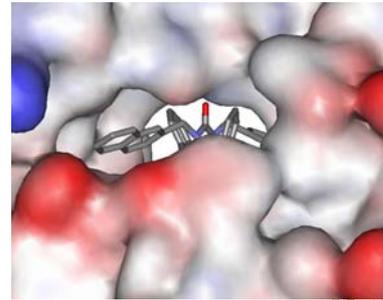
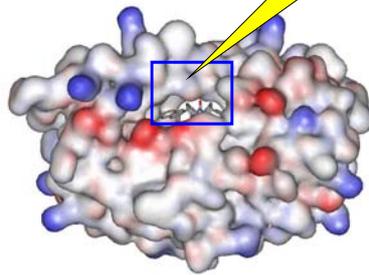
コンピュータを使った構造解析による 化合物の探索

- 抗エイズ薬



XK263

このポケットに
フィットする構造
の化合物を探索



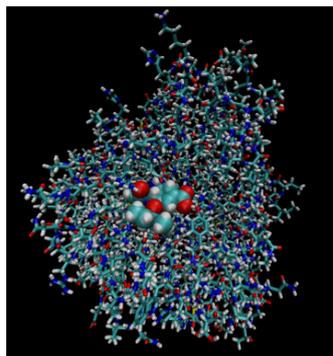
31

bioGrid

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。



インフルエンザの原因タンパク質に結合するタミフル

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

32

bioGrid

**「京」産業利用枠:新薬開発を加速する「京」インシリコ創薬基盤の構築
KBDDコンソーシアム“K” supercomputer-Based Drug Discovery**



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 奥野恭史

製薬企業(22社):

旭化成ファーマ(株)、アステラス製薬(株)、アスピオファーマ(株)、エーザイ(株)、小野薬品工業(株)、科研製薬(株)、(株)カネカ、キッ
セイ薬品工業(株)、杏林製薬(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)、帝人ファーマ
(株)、東レ(株)、日産化学工業(株)、日本新薬(株)、富士フイルム(株)、マルホ(株)、MeijiSeikaファルマ(株)、持田製薬(株)

IT企業(2社): (株)京都コンステラ・テクノロジーズ、三井情報(株)

大学等: 京大医学部、横浜市大、兵庫県大

(公財)都市活力研究所、産業技術総合研究所、理研、(公財)先端医療振興財団

- 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うる計算フロー(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

プロジェクトの概要

(1) 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する

利用アプリ: CGBVS(京大・医・奥野)

Chemical Genomics Based Virtual Screening



(2) 上記のうち蛋白質10×化合物35程度につ
いては結合自由エネルギーを計算する

利用アプリ: MP-CAFEE(東大・生産研・藤谷)

Massively Parallel Computation of Absolute binding Free Energy



34

プロジェクトのねらい

(1) アプリケーションの産業利用可能性の検証

(2) ワークフローの構築

(3) 製薬企業の「京」利用スキルの習熟

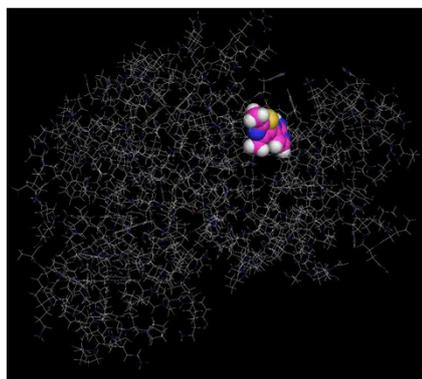


(4) スパコン利用者のコミュニティの形成

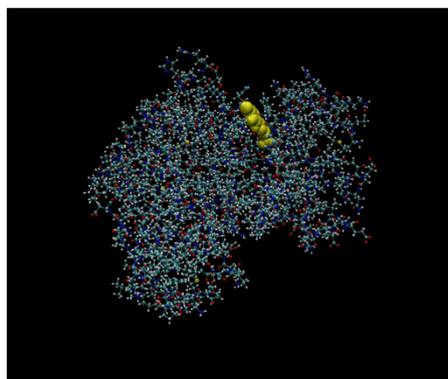


35

計算結果



PCクラスターによる計算
形状のみによるドッキング: 精度低い



「京」による計算
一定時間の動きを解析: 精度高い



36

実際の処理



37

Multiple Target Screening

1つの蛋白質に対して、どの化合物が結合しやすいかは、予測困難。
 1つの化合物に着目すると、どの蛋白質により結合しやすいかは、かなり正確にわかる。

結合エネルギーを表示

化合物ライブラリー(chemical compound library)

	1	2	3	4	5	6	7	8	9	10	11
1	50	90	8	3	60	50	4	2	4	40	6
2	9	60	9	2	2	6	5	3	3	5	5
3	10	80	10	12	3	5	6	5	2	4	4
4	11	90	5	1	6	4	2	1	1	3	3
5	4	50	7	11	5	3	7	6	11	2	2
6	3	80	6	4	4	12	1	4	10	6	1
7	8	90	2	5	9	2	8	7	9	7	12
8	12	60	11	6	8	11	9	8	8	8	11
9	2	80	3	10	7	10	10	12	12	9	10
10	7	70	12	7	12	9	11	9	7	10	9
11	6	60	4	8	10	8	12	10	6	11	8

標的蛋白質

HIT化合物?

どの蛋白質にも結合している。

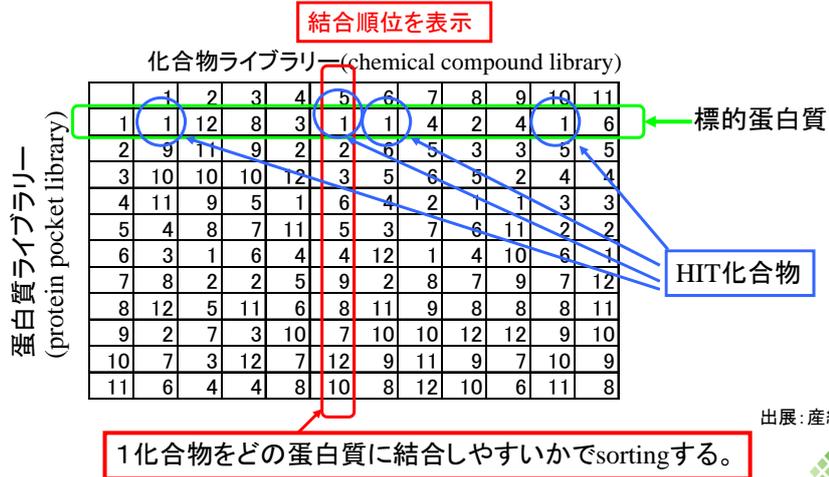
出展:産総研 福西

38



Multiple Target Screening

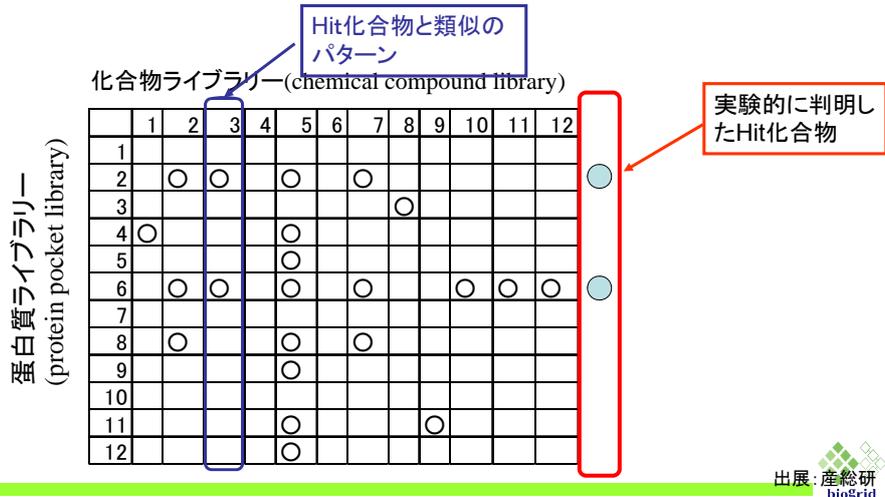
1つの蛋白質に対して、どの化合物が結合しやすいかは、予測困難。
1つの化合物に着目すると、どの蛋白質により結合しやすいかは、かなり正確にわかる。



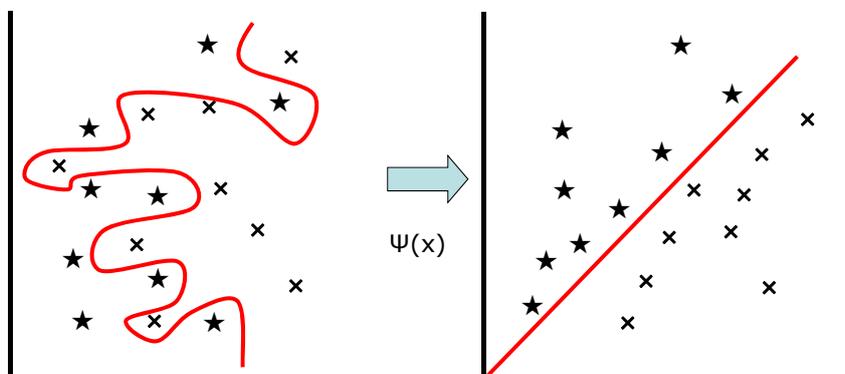
Docking Score Index

Dockingシミュレーションを利用した、化合物の特徴抽出

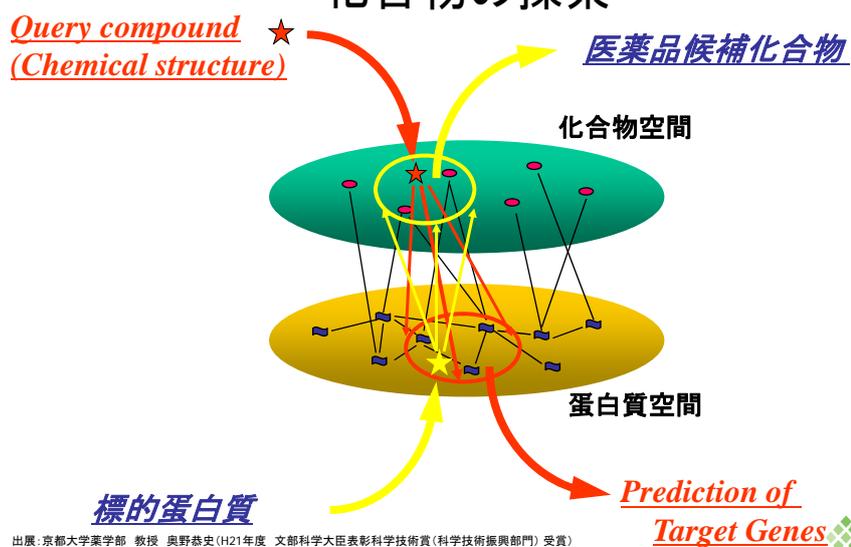
化合物の各種蛋白質へのドッキングスコアの標準的パターンを抽出する。



統計解析的アプローチの原理

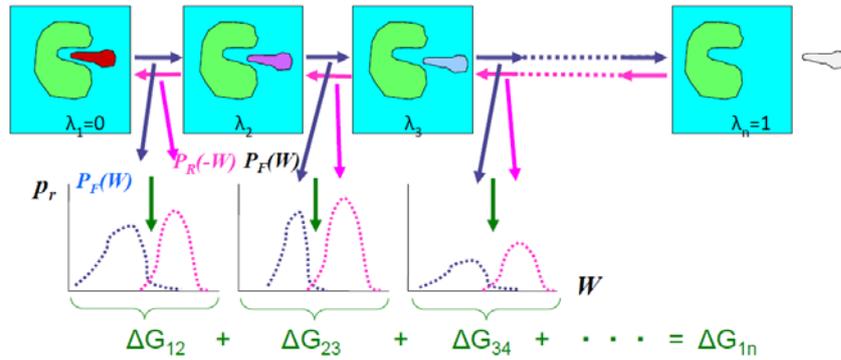


コンピュータを使った統計解析による 化合物の探索



Jarzynskiの式による結合自由エネルギーの算出

Annihilate interaction strength of ligands to other molecules: λ_i



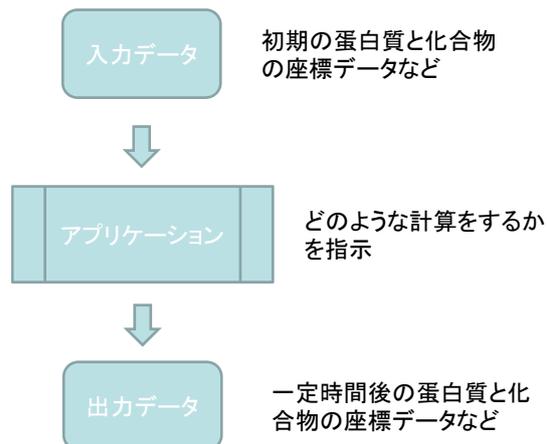
Interaction strength Trajectories Total
 32 λ points X 12 = **384**

出展: バイオグリッド勉強会資料より

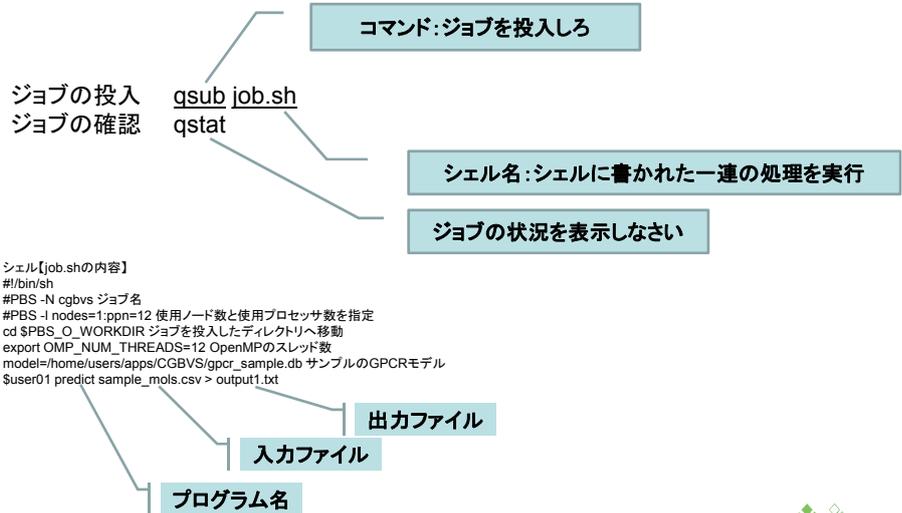
藤谷らの方法

FKBPは12初速度モデルだったが、我々は計算時間積節約のために6初速度モデルとした

スパコンの起動の実際



起動コマンドの実際



GUIツールの開発

(スパコンをマウスで操作できる)

ジョブ名	ユーザー	ノード数	プロセッサ数	開始日時	終了日時	ステータス	計算時間
114.176.118.218	5428	12	12	2015/11/26 00:45:13	2016/02/24 12:21:18	実行中	
114.176.118.218	5428	12	12	2015/11/26 00:45:13	2016/02/24 12:21:18	実行中	
114.176.118.218	5428	12	12	2015/11/26 00:45:13	2016/02/24 12:21:18	実行中	



創薬と人工知能

2016.11.17 LINC(Life Intelligence Consortium)を発足
製薬など43、IT系企業28、アカデミア(京大、理研、医薬基盤研など)が参加

- 人工知能とは

- 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

- 具体例

- 蛋白質と化合物の結合を学習
⇒新たな蛋白質に結合する化合物を予測
 - アミノ酸配列と蛋白質の構造を学習
⇒新たなアミノ酸配列の蛋白質の構造を予測



47