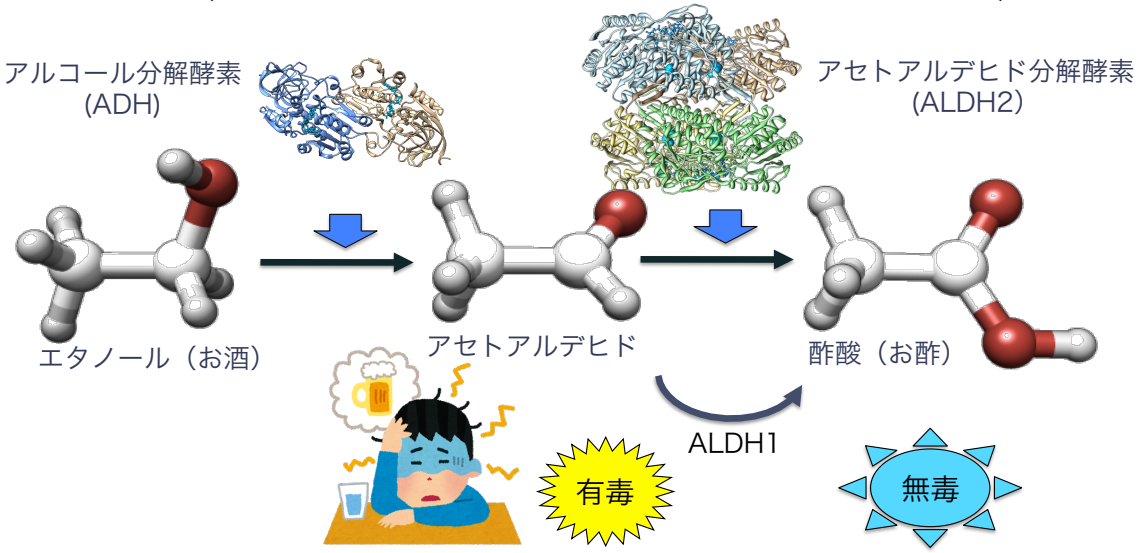
# ゲノミクスからの構造インフォマティクス

**白井　剛**
**(長浜バイオ大)**

1) 疾患ゲノミクスの解析には、タンパク質複合体の立体構造情報が必要である。
2) タンパク質複合体の立体構造解明には、構造インフォマティクスが必要である。
3) 構造インフォマティクスでどのような事がわかるか？
4) 現在の構造インフォマティクスには何が欠けているか？

## 1 塩基多型(SNP)とお酒の強さ

1) お酒(エタノール)は体内で、アルコール分解酵素によってアセトアルデヒドに、アセトアルデヒド分解酵素(ALDH2)によって酢酸に代謝される。
2) アジア人の多くは不活性なALDH2の1塩基変異対立遺伝子(SNPアリル)を持つ。これによりアセトアルデヒドを迅速に代謝できず、有害なアセトアルデヒドが体内に蓄積しアルコール中毒を起こす(ただし代謝は遅いものの、別の酵素ALDH1による迂回反応がある)。



# 変異ALDH2の代謝効率は立体構造により説明される

1) 変異型ALDH2は、１箇所のG(グアニン)がA(アデニン)に変異したSNPであり、タンパク質ではサブユニット界面近くのグルタミン酸(E)がリシン(K)に変異する。

2) 父方/母方の遺伝子がG/Gの場合の活性を１とするとA/Aでは０(不活性)になる。しかしG/Aでは半分ではなく1/16になる。これはALDH2が４つのタンパク質(サブユニット)が複合体を作って働くので、完全な複合体は$(1/2)^4=1/16$になるため。

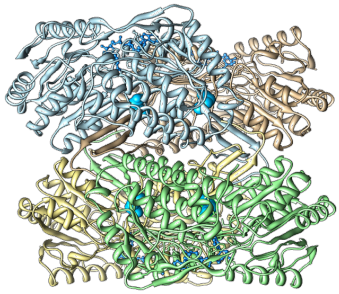
Gタイプ

```
...ggcagccattactcgtcctcactcccacaccaacaacctc
catccagtgcctgccGcagccgcttctgctgcagcggggacgc
gtgcaagtacaggaggatatccgcttccattactgcgctgcgc
cgcggcggaaacagcagcagcagagggg...
```

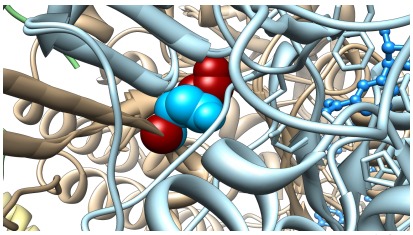
Aタイプ

```
...ggcagccattactcgtcctcactcccacaccaacaacctcc
atccagtgcctgccAcagccgcttctgctgcagcggggacgcgt
gcaagtacaggaggatatccgcttccattactgcgctgcgccgc
ggcggaaacagcagcagcagagggg...
```

ALDH2　野生型E487

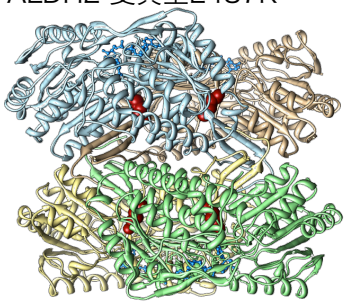


E487K



PDB 1o02

ALDH2 変異型E487K

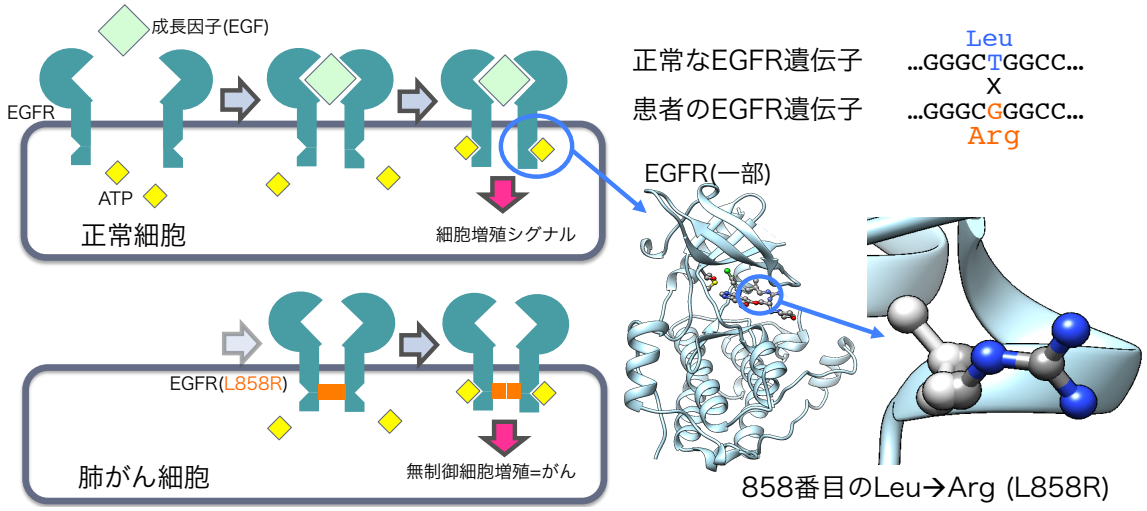


| 遺伝子型 | G/G | G/A | A/A |
|---|---|---|---|
| 活性 | 1 | 1/16 | 0 |

# なぜ病気になるのか？

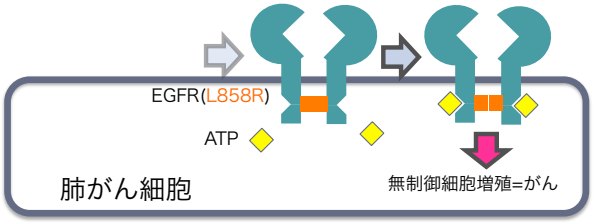
1) 細胞の増殖は制御されている。成長因子(ホルモン)が分泌される→細胞表面の上皮成長因子受容体(EGFR：タンパク質)に結合→EGFRが2量体化→細胞内でATPを結合しリン酸化される過程を経て、細胞核に増殖(分裂)シグナルが伝達される。

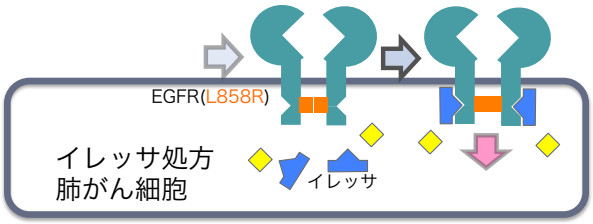
2) 肺がん患者の多くで、EGFRの858番目のアミノ酸Leu(ロイシン)がArg(アルギニン)に変異している(L858R)。L858Rは成長因子がなくてもEGFRが2量体化し、無制限に増殖シグナルを発生する。つまり、窒素原子が差し引き3原子増えるだけで病気になる！

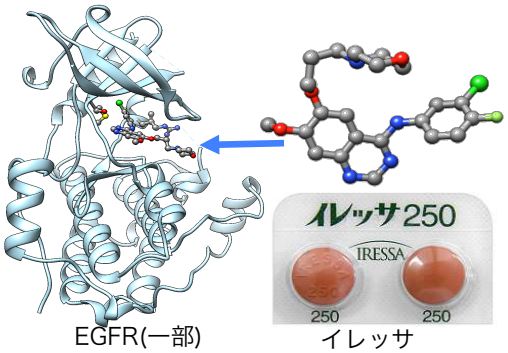


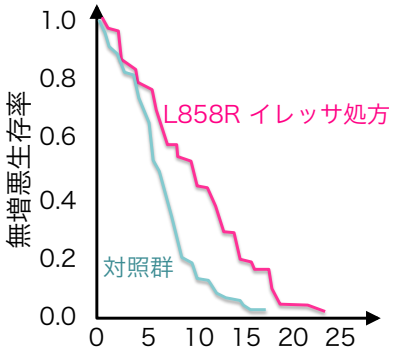
# なぜ薬で病気が治るのか？

1) EGFRはATPを結合・分解しないと増殖シグナルを送れない。これを特異的にブロック(阻害)すればがん細胞の増殖を止められる。
2) 抗肺がん薬イレッサは、 L858RのEGFRのATP結合部位に特異的に結合し、ATPの結合を阻害する。



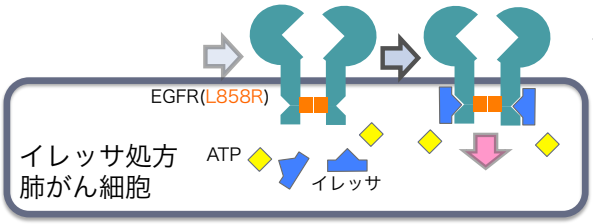



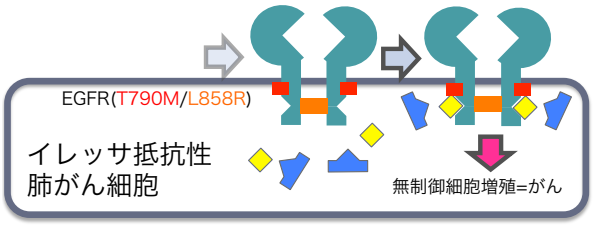




Mok *et al. New Engl J Med.* **361**, 947(2009)

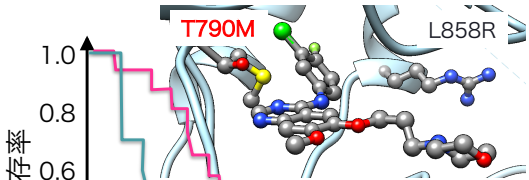
# なぜ薬が効かないのか？

1) ところが、L858Rに加えてEGFRの790番目のアミノ酸Thr(トレオニン)がMet(メチオニン)に変異(T790M/L858R)するとイレッサが結合しにくくなり、再び無制御状態で増殖シグナルを発生する。
2) これが、抗がん剤や抗ウイルス剤を服用し続けると現れる薬剤耐性の原因の一つである。薬剤という選択圧を与えることで、耐性細胞や耐性ウイルスの細胞内「進化」を助けてしまう。




イレッサ有効EGFR遺伝子 …ATCA**C**GCAG…
(Thr)
X
イレッサ耐性EGFR遺伝子 …ATCA**T**GCAG…
(Met)



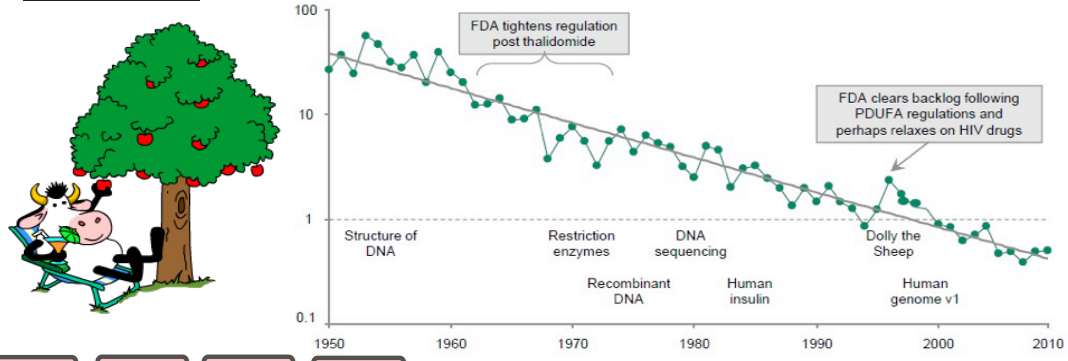




Shyamala *et al., N Engl J Med* 359, 366 (2008)

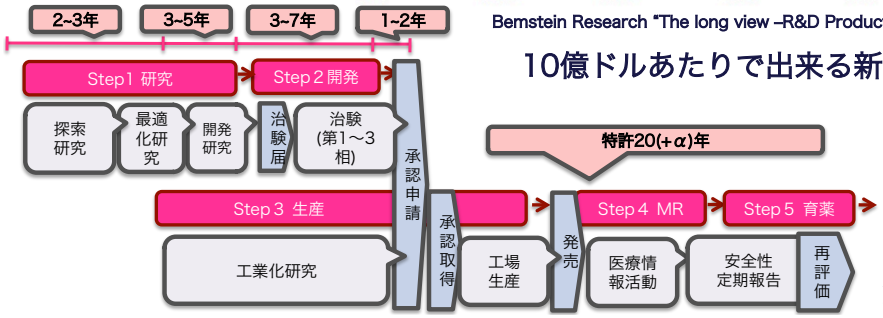
# Low hanging-fruit (容易なドラッグターゲット)の枯渇

1) 新薬研究開発費は高騰し続けており、探索研究のアウトソーシング/オープンイノベーション化が必要とされる。




Bemstein Research "The long view –R&D Productivity" (Sep. 30 2010

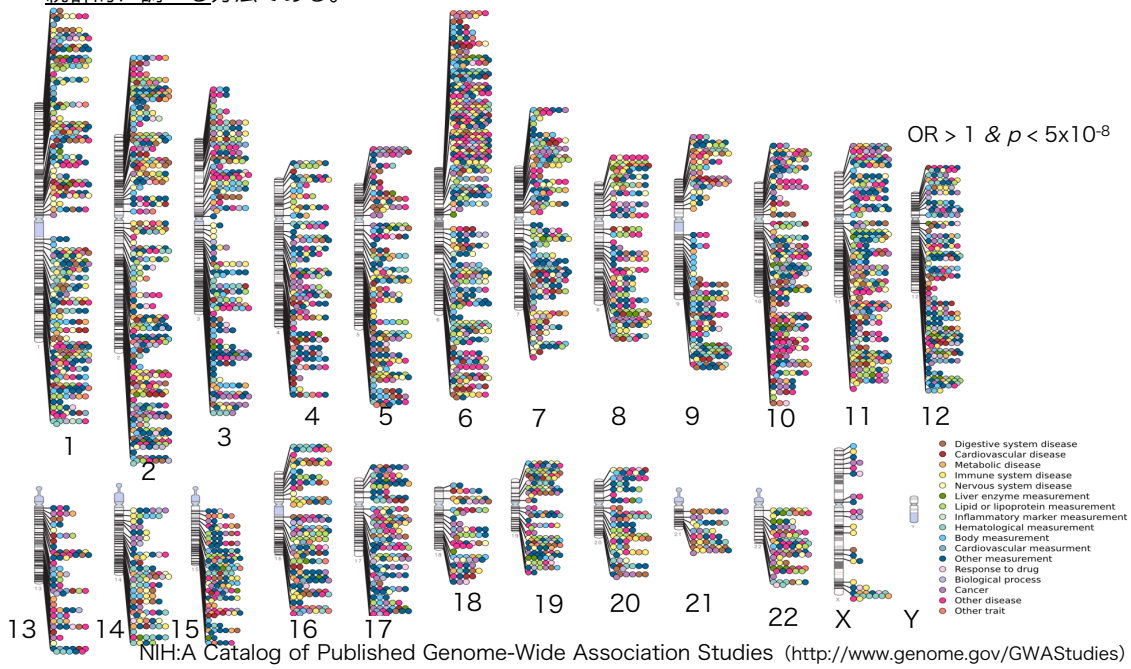
10億ドルあたりで出来る新薬の数




創薬プロセス

# 全ゲノム相関解析(GWAS = Genome Wide Association Study)

1) GWASとは、ゲノム全体をカバーする多数の１塩基多型（SNP）や繰り返し配列多型の遺伝子型を多人数の被験者(コントロールを含む)について決定し、主に多型の頻度と疾患や量的形質との関連を統計的に調べる方法である。

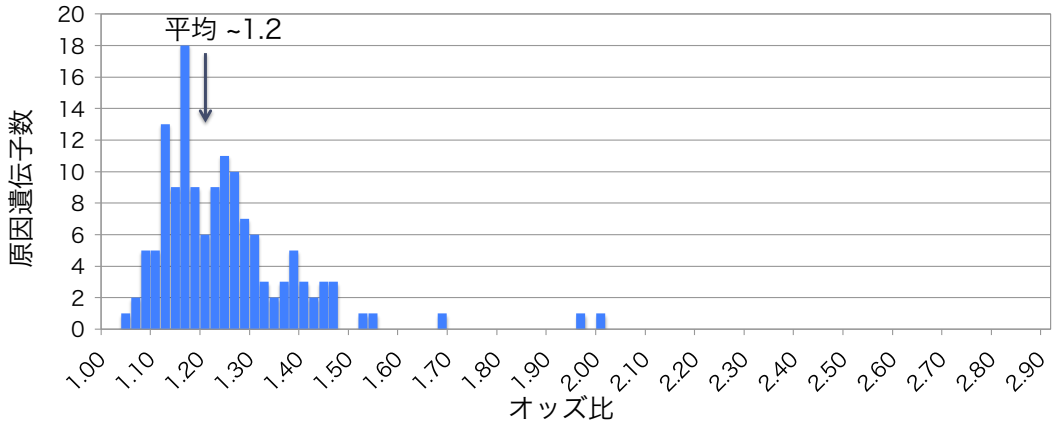



NIH:A Catalog of Published Genome-Wide Association Studies（http://www.genome.gov/GWAStudies）

# よくある病気(Common diseases)のオッズ比

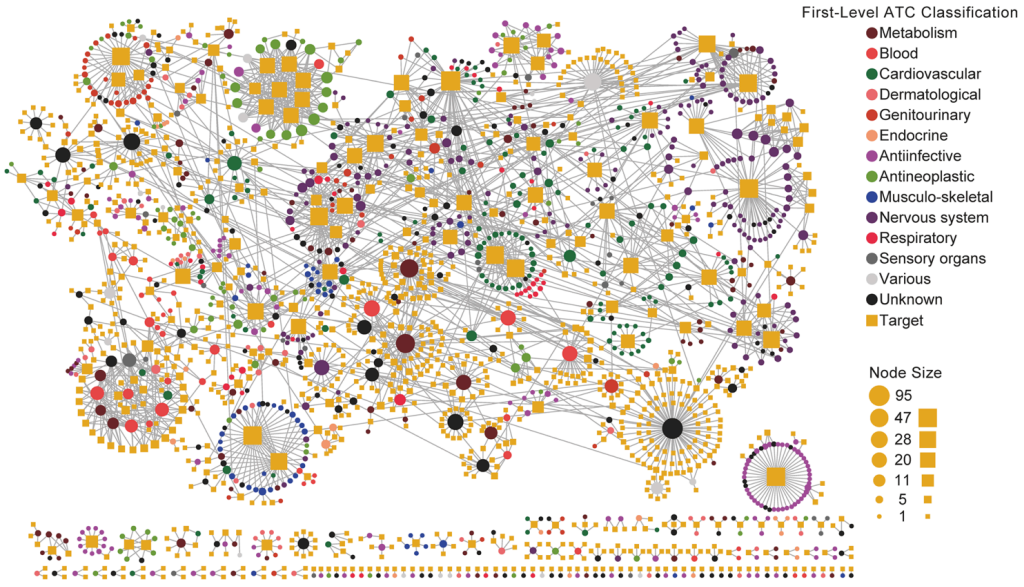
1) 下のグラフは、日本人の３大疾病である癌、心筋梗塞、脳卒中(あわせて日本人の~53%の死因)、および国民病とも呼ばれる糖尿病について行われたGWASによって発見された関連遺伝子変異のオッズ比(罹患リスク)分布を示したものである。

2) オッズ比の平均は1.23に過ぎない。また２を越えるオッズ比を示す原因遺伝子が見つかることは極めてまれ。これは単一の遺伝子の変異で説明できる病因はほとんど見つからないことを示している。



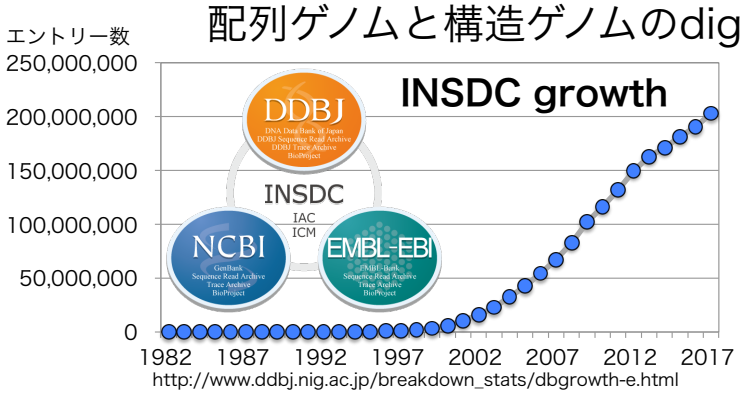

# Drug–Target Protein network (DTP network)

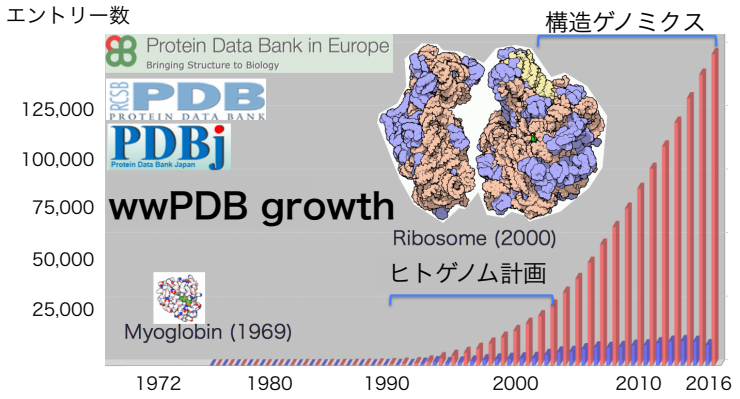
1) 創薬ターゲット(=病気の原因)は遺伝子/タンパク質ネットワーク、すなわちPPI (Protein-Protein Interaction/Interface)である。




Muhammed et al., *Nature Biotechnology* 25, 1119-1126 (2007)

# 配列ゲノムと構造ゲノムのdigital-divide




1) 配列データサイズと構造データサイズの乖離は加速している。
2) ゲノム(配列)データはすでに多くの生物について”complete”であるのに対して、構造データはすべての生物について”in-complete” である。
3) このギャプを埋めることが構造インフォマティクス(による分子モデリング)の主要な役割の1つである。

https://pdbj.org/info/statistics

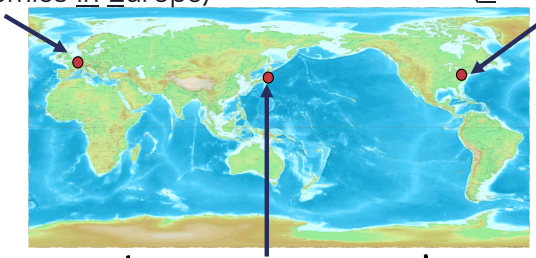
# 構造ゲノミクス(structural genomics)

1) 2004年に解読宣言のあったヒトゲノムに代表される、全遺伝子配列に対して解析・研究を行う立場をゲノミクスという。
2) 構造ゲノミクスとは、ポスト(後)ゲノミクスの課題として、「代表タンパク質」の立体構造を解析・研究することをいう。構造ゲノミクスに対して1)を配列ゲノミクスという場合もある。

Human Genome Project 1990-2003

EU : SPINE 2002-2006
(Structural Proteomics In Europe)

USA :PSI1 ⇨ PSI2 ⇨ PSI3
(Protein Structure Initiative)
2000-2005
2005-2010
2010-2015

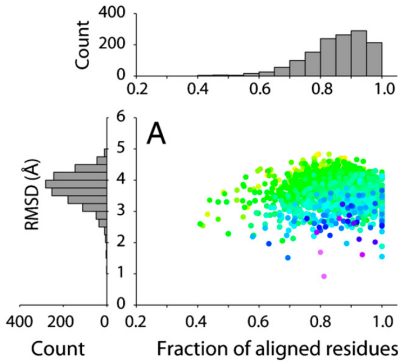


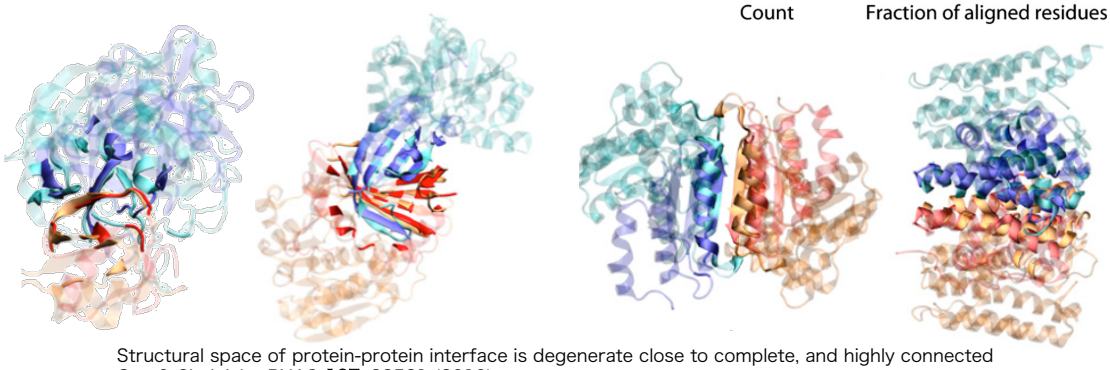
日本:タンパク3000 ⇨ ターゲットタンパク ⇨ 創薬等基盤プラットフォーム
2002-2006 2007-2011 PDIS 2012-2016 BINDS 2017-

複合体構造ゲノミクスの必要性 **PSI Has to Live and Become PCI: Protein Complex Initiative**
Ilya A. Vakser Structure *16*, January 2008

# PPIインタフェース構造の情報は既に十分存在する(?)

1) PDBのPPI構造を分類すると90%程度はいずれかの既知インターフェース(分子間の接触面)に類似している(残基比較で~3.5Å　RMSD，~0.75 coverage程度)。PPIインターフェース構造は～1000種類に分類可能である。
2) 80%程度のPPIインターフェース構造は、7種類の代表的構造に関係づけられる。

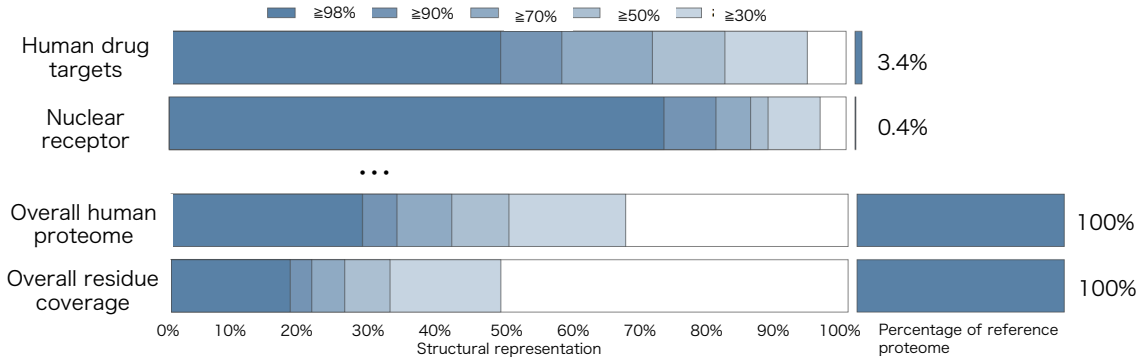


Structural space of protein-protein interface is degenerate close to complete, and highly connected
Gao & Skolnick., *PNAS*, **107**, 22521 (2010)

# 創薬ターゲットの立体構造は既に十分存在する(?)

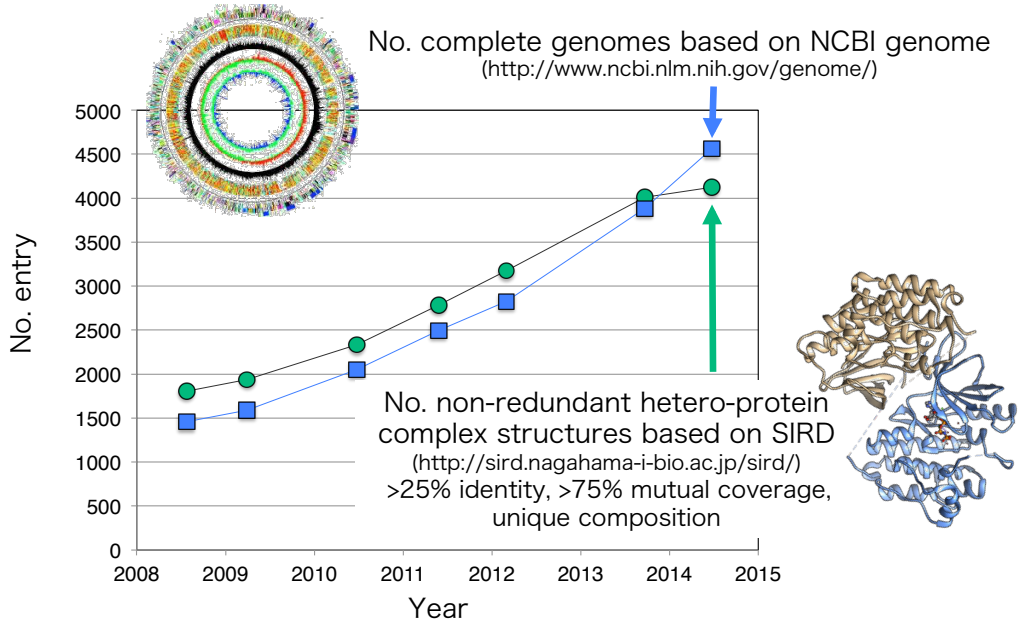
1) ヒト遺伝子(~21,000)のうち、694(3.4%)遺伝子/タンパク質がドラッグターゲットである。
2) ヒトタンパク質の70%(遺伝子数)が構造既知(≧30%のアミノ酸配列一致度を示す既知構造がある)であり、ドラッグターゲットに限ると95%の構造がすでに解明されている。




Structural coverage of the proteome for pharmaceutical application
Somody *et al. Drug Discovery Today* (2017)

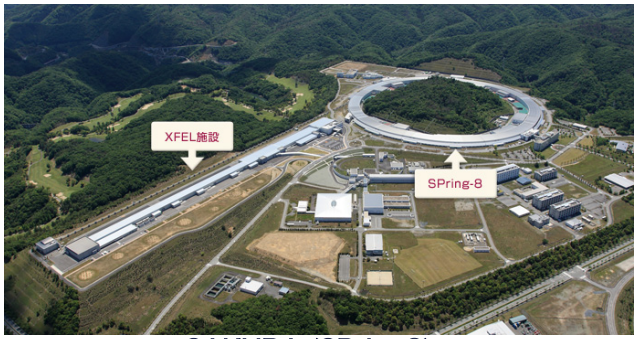
# ゲノム生物学 *vs* 複合体構造ゲノミクス

1) 配列データベース中の既知ゲノム数と構造データベース中の既知ヘテロ複合体構造数は同じくらいである。

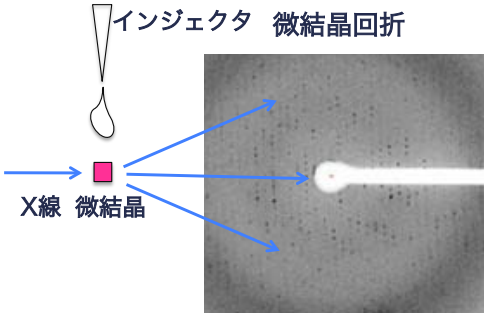


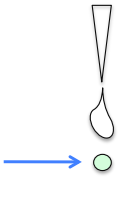

# X線自由電子レーザー(XFEL)による微結晶/単粒子X線回折

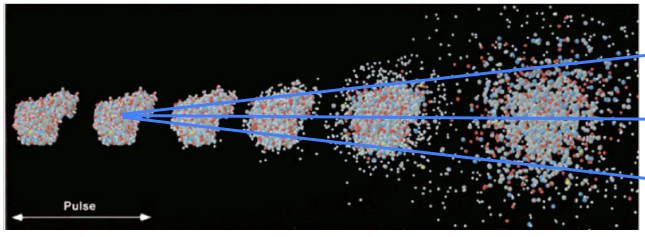
1) 高輝度・高コヒーレントX線によって、微結晶X線回折実験、さらには単粒子X線回折実験が可能になる。
2) 1サンプル(分子・結晶)1照射なので、X線損傷がなく時間分割解析に適している。




SAKURA (SPring8) http://xfel.riken.jp/

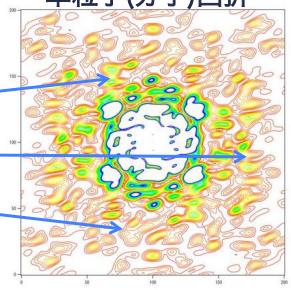


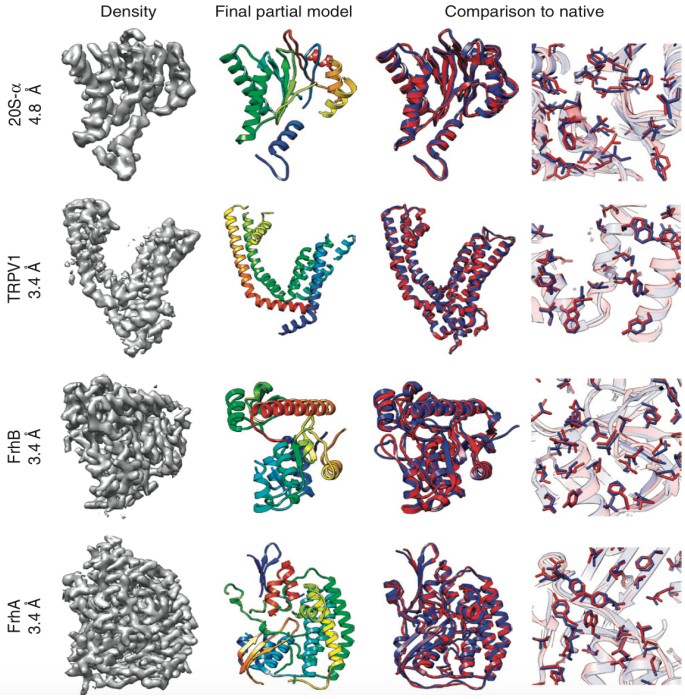

クーロン爆発

単粒子(分子)回折



# 原子分解能クライオ電子顕微鏡単粒子解析

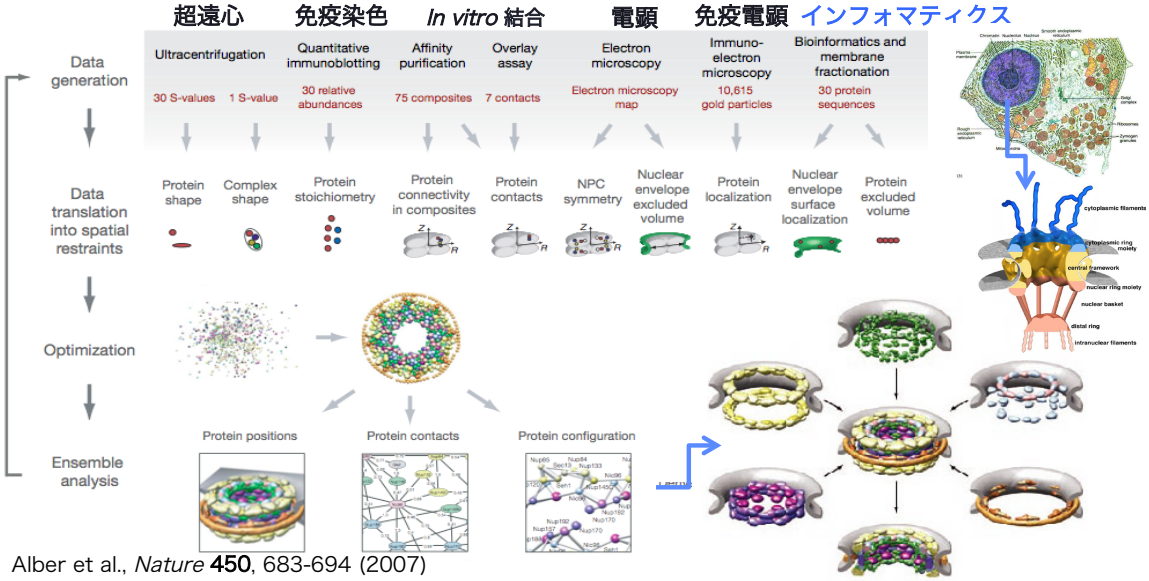
1) コンピューター制御される高性能のステージ(多試料搭載・交換可能)を搭載した専用のクライオ電子顕微鏡の開発，CMOS電子直接検出器(直接カウントによる画像ノイズ・劣化軽減)，新規アルゴリズム(Relion　ベイズ統計による電顕画像解析)による解析ソフトウェアの登場により、準原子分解能(near-atomic resolution: 2~3A分解能)での低温電顕単粒子解析(cryo-EM single particle analysis)が可能になった。
2) 2~3A分解能ではタンパク質の2次構造を明確に識別可能であり、場合によっては側鎖のモデリングもできる。事実上X線結晶解析に匹敵する精度の分子構造が得られる。




Wang *et al*., *Nature Method*, **12**, 225 (2015)

# 「あらゆる方法を使う」方法

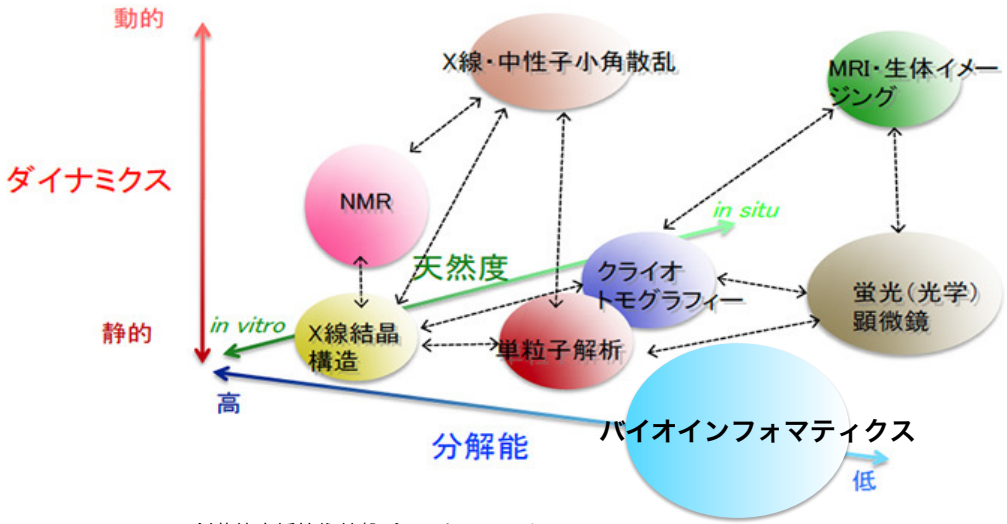
1) 核膜孔複合体は核<->細胞質の高分子輸送を制御し、細胞内のタンパク質局在の管理に重要。分子量120MDa、約30の異なるタンパク質からなる超分子で、実験的構造解析は極めて困難である。
2) そこで、実験(超遠心..結合**アッセイ..構造解析)から計算機解析(**インフォマティクス)までを統合して、一つの複合体構造を決定(予測)する。
3) インフォマティクスの役割は、サブユニットの構造予測とサブユニットドッキング(PPI)予測である。




Alber et al., *Nature* **450**, 683-694 (2007)

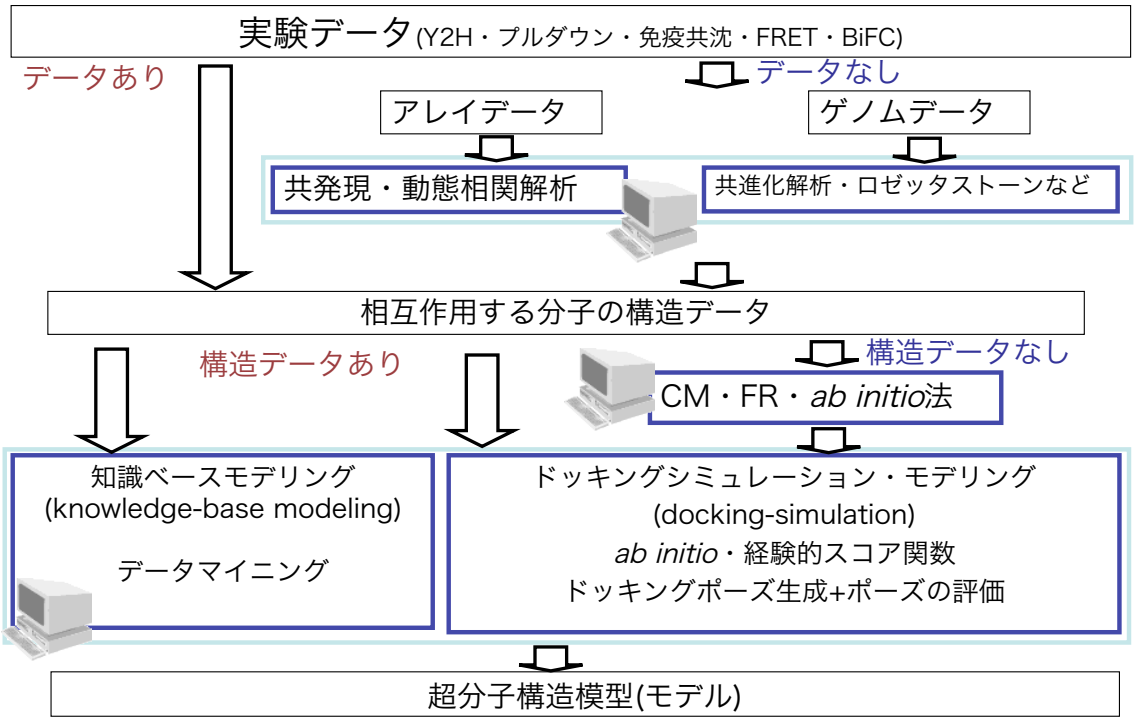
# 相関構造解析(correlated structure analysis)

1) 様々な実験＋理論手法からのデータを総合して生体超分子の構造・ダイナミクス・機能を解析する研究を、相関構造解析(correlated structure analysis または Hybrid/Integrated structure analysis)という。

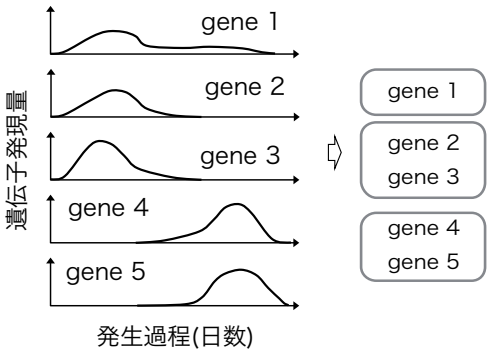



創薬等支援技術基盤プラットフォーム　http://pford.jp/p4d/sac1/analysis.html

## PPIモデリング:フローチャート

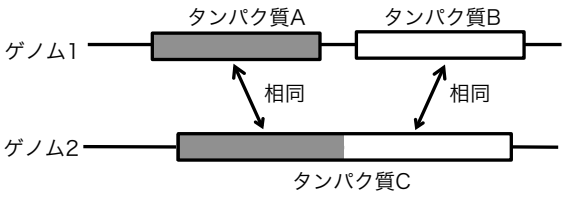


# 相互作用するタンパク質を予測する方法




1) 共発現・動態相関解析

マイクロアレイデータなどを使って、遺伝子発現の増減が組織・発生時期・細胞応答などに対して相関する(挙動を共にする)遺伝子を探索する方法。高い相関を持つ場合にタンパク質レベルで相互作用していると推定される。

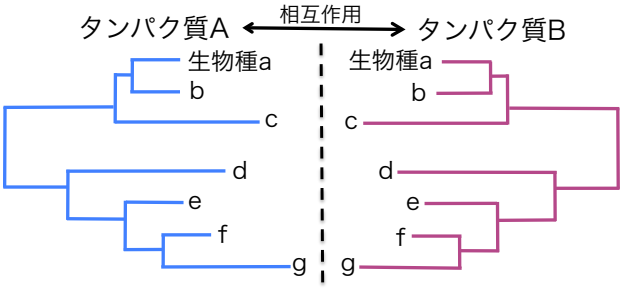


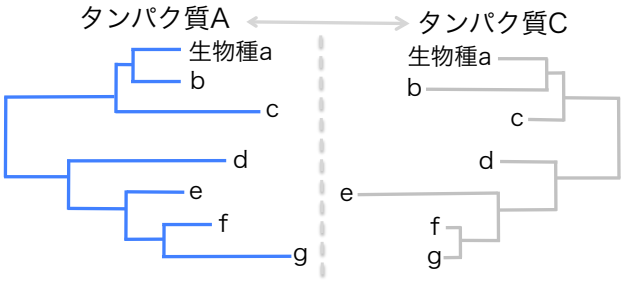

2)ロゼッタストーン

複合体を形成するタンパク質の遺伝子は、進化の過程を通じて共存する必要がある。様々な生物のゲノムを比較して、共存確率の高いタンパク質を予測する方法。

特に左図のように、ある生物(ゲノム1）では別の遺伝子にコードされているが、他の生物(ゲノム2)では1つのタンパク質のドメインとして存在するような関係は、タンパク質AとBが同時に発現し同じように局在する必然を示唆するので、ロゼッタストーンと呼ばれ、相互作用の有力な証拠とみなす。

# 相互作用するタンパク質を予測する方法



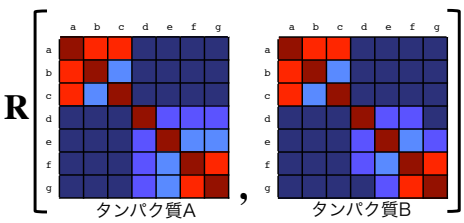




3)共進化解析・ミラーツリー法

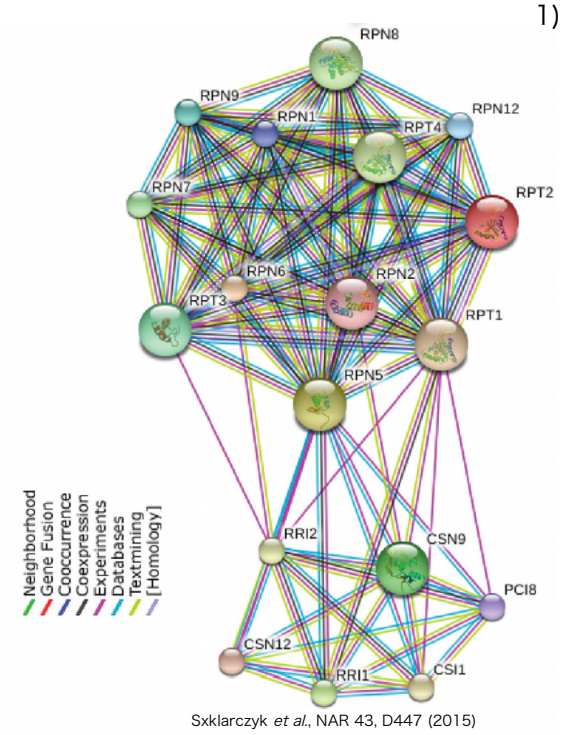
複合体を形成するタンパク質の遺伝子は、類似した進化過程を経る傾向がある。

典型的な例としては、分子系統樹のある枝でタンパク質Aの進化速度が加速/減速している場合に、相互作用するタンパク質Bの進化速度も加速/減速する。したがって系統樹の樹形は類似する(系統樹が鏡に写した関係になる=ミラーツリー)。

系統樹の類似性は、もっとも簡単には生物種間の進化距離マトリクスの相関係数を求める。



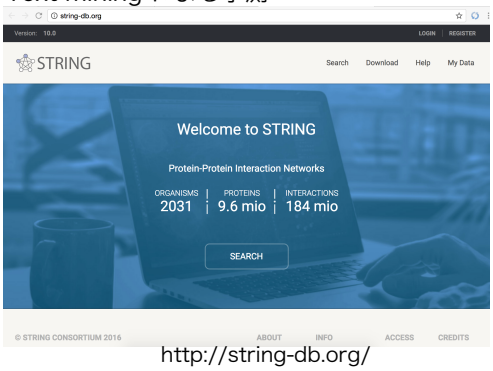
# PPI予測データベース




Sxklarczyk *et al.*, NAR 43, D447 (2015)

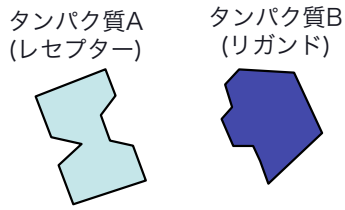
1) STRINGは共進化解析・ロゼッタストーンによるPPI予測の総合的なデータベースであり、以下のバイオインフォティクス予測および実験データ、文献データが集積されている。

a) Neighborhood ゲノム上の局在による予測(主に原核生物のオペロンのように機能的にk関連した遺伝子がゲノム上でクラスタする傾向を利用する予測法)
b) Gene Fusion 遺伝子融合(ロゼッタストーン)による予測
c) Cooccurrence ゲノムの共存在による予測
d) Coexpression 共発現による予測
e) Text mining による予測

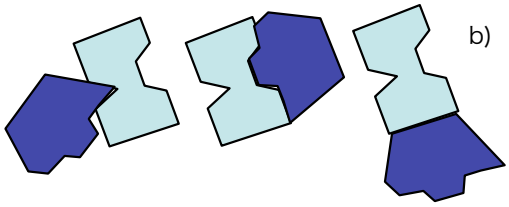



http://string-db.org/

# ドッキングポーズを予測する方法




複合体構造(ドッキングポーズ)

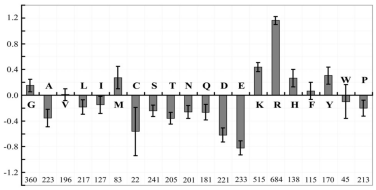


...NWVVTAAHCGVTT...LTINNDITLLKL...
...NYLVTGAHAGVST...V---NDVTVFKH...
...NWVVTAAHCALSS...LTIQNDIVLLRV...
...AICFAVYHG-GSR...---EQDLVAWHG...
...GVCWALYHC-GSK...---DQDLIGWPA...

保存残基(マルチプルアラインメント)

1) ドッキングポーズの生成
結合する一方のタンパク質をレセプター、他方をリガンドとした時、レセプターに対してリガンドの相対配置をドッキングポーズという。
FFT法(表面の静電ポテンシャルや疎水性などの性質をフーリエ変換することで、高速に相関をとる)やハッシュ法(水素結合や静電相互作用などの安定化相互作用部位の相対配置のリストをあらかじめ作成し、高速にマッチングする)などの高速探索法を用いて、多数のドッキングポーズを生成する。

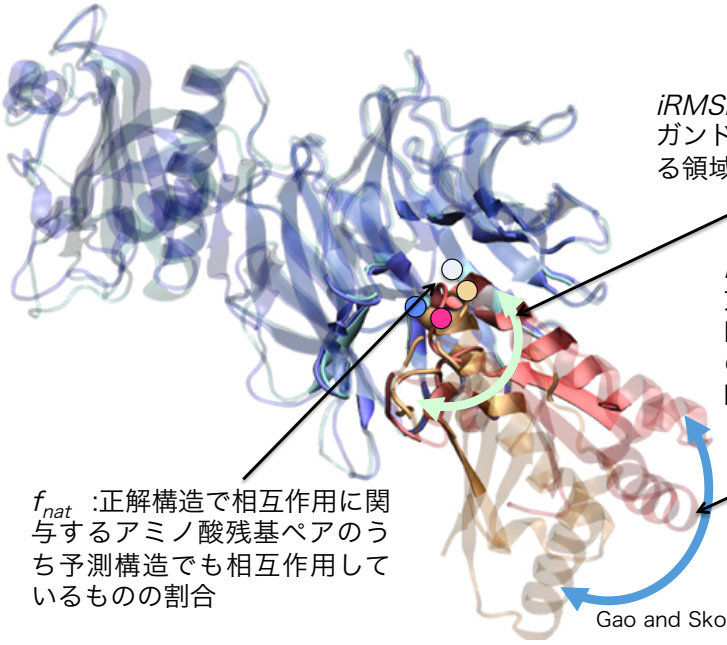
2) ドッキングポーズの評価
関数を設定してドッキングポーズの安定性を評価する。
*a) ab initio*法　分子(動)力学計算に用いるのと同様、あるいは簡素化されたエネルギー関数を用いて評価する
b) 経験的方法　相互作用部位が進化の過程で保存される傾向を利用して保存傾向を評価する方法、あるいはアミノ酸ごとに相対的に相互作用部位に関与する傾向値[もっとも簡単には、対数オッズlog (アミノ酸*i*が相互作用部位に観察される相対頻度)/(同アミノ酸が分子表面一般に観察される相対頻度)]をスコアとするなど様々な経験的評価方がある。




アミノ酸の相互作用部位傾向値

# タンパク質ドッキング精度の評価基準(CAPRI)

1) ドッッキングの結果(ドッキングポーズ)の評価には、相互作用に関与する残基ペアの一致率($f_{nat}$)，リガンド側の予測-正解の主鎖rmsd(*IRMSD*)、相互作用に関与する残基のみの主鎖rmsd(*iRMSD*)が使われる(ドッキングコンテストCAPRIの基準)。



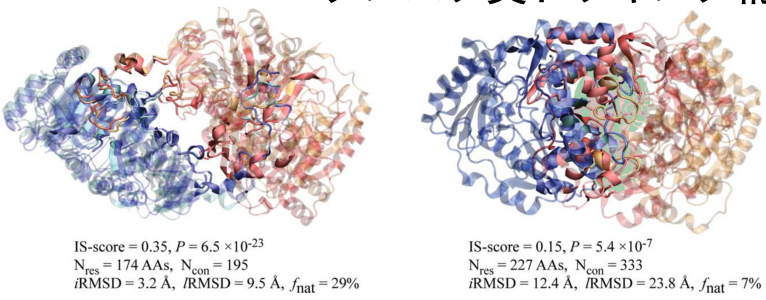
*iRMSD*: 正解-予測レセプター/リガンド主鎖間の相互作用に関与する領域のみのCαのRMSD (Å)

*IRMSD*: レセプター(通常大きい方のタンパク質)を重ね合わせた時の正解-予測リガンド(小さい方のタンパク質) 主鎖間のCαの正RMSD (Å)

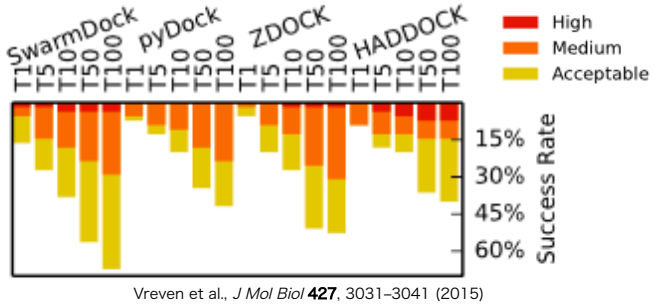
$f_{nat}$ :正解構造で相互作用に関与するアミノ酸残基ペアのうち予測構造でも相互作用しているものの割合

Gao and Skolnik, Proteins. 79, 1623–1634 (2011)

# タンパク質ドッキング精度




Gao and Skolnik., Proteins **79**, 1623 (2011)



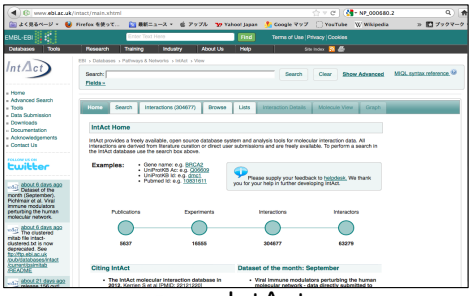

Vreven et al., *J Mol Biol* **427**, 3031–3041 (2015)

1) $F_{nat}$, *IRMSD*, *iRMSD*によりドッキングポーズはhigh, medium, acceptable（通常ここまでが正解）, incorrectに分類される。
2) 公開ドッキングソフトでよく利用されるSwarmDock, pyDock, ZDOCK, HADDOCKなどのベンチマークテストの結果から、トップのポーズ(T1)で15%前後、トップ10のポーズ(T10)で20~40%、トップ100のポーズ(T100)で40~60%程度の確率で、すくなくとも1個のacceptable以上の解が得られる。

High: $f_{nat} \geq 0.5$ & (*IRMSD* ≤ 1Å || *iRMSD* ≤ 1Å)
Medium: ($f_{nat} \geq 0.5$ & *IRMSD* > 1Å & *iRMSD* > 1Å) || ($f_{nat} \geq 0.3$ & $f_{nat} < 0.5$ & *IRMSD* ≤ 5Å & iRMSD ≤ 2Å)
Acceptable: ($f_{nat} \geq 0.3$ & *IRMSD* > 5Å & *iRMSD* > 2Å) || ($f_{nat} \geq 0.1$ & $f_{nat} < 0.3$ & *IRMSD* ≤ 10Å & *iRMSD* ≤ 4Å)
Incorrect: $f_{nat} < 0.1$ || (*IRMSD* > 10Å & *iRMSD* > 4Å)

# タンパク質複合体の知識ベースモデリング

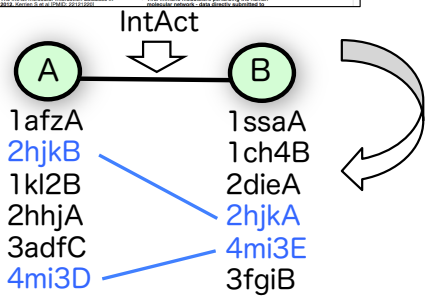
1) 相互作用データベース(IntAct)とタンパク質構造データベース(PDB)を相関させて、「すでに潜在的に知っている(はずの)」複合体構造の知識ベースモデルを構築する。



Protein-Protein Interaction Database

Contents as of Oct. 2014

| | |
|---|---|
| Number of Interactions: | 460,871 |
| Number of Proteins: | 77,009 |
| Number of Experiment: | 34,785 |





Protein Structure Database PDB

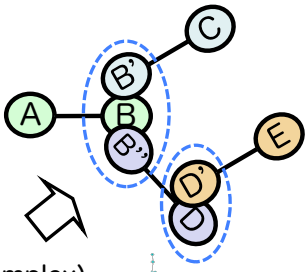
Contents as of Jun. 2014

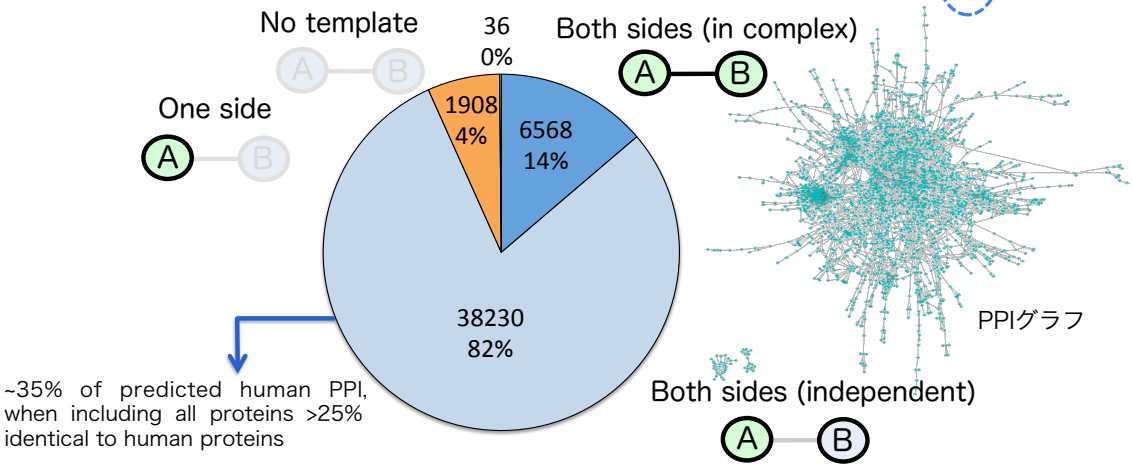
| | |
|---|---|
| Number of structures: | 93,900 |
| Number of subunits: | 241,805 |

Tsuji *et al.*, *Scientific Reports* 5:16341 (2015)

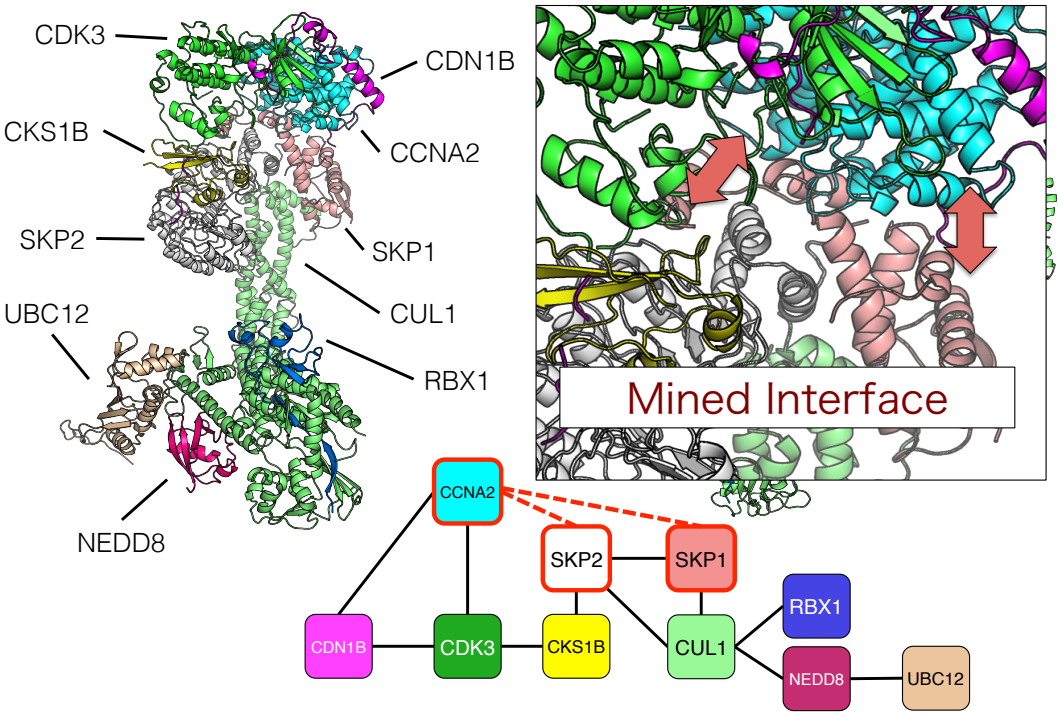
## PPI (Protein-Protein Interaction)から複合体モデルを構築する

1) 残念ながら我々の知識は、ヒトのタンパク質複合体(2体間相互作用ベース)の5~14%にすぎない。
2) しかし、2024個のタンパク質からなる巨大相互作用ネットワークの構造をすでに「知って」いる。これを組み合わせることで、より大きな複合体の構造を予測できる。



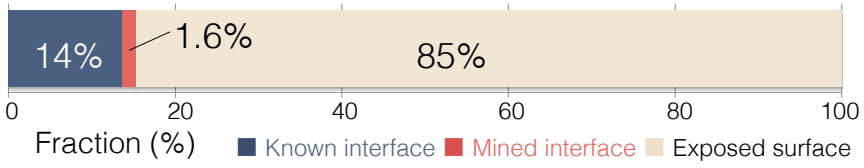
# Cyclin - Ubiquitin ligase complex model




## Disease-related variants on mined interface

1) ~3,200個のヒト超分子モデルが構築可能で、～1,300のモデルに新規インターフェースが存在する。
2) 44個の疾患関連変異が新規インタフェースにマッピングされる。そのうち10疾患については、PPIインターフェース変異の関与が初めて指摘されるケースである。

Molecular surface properties of supramolecular models



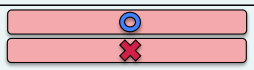
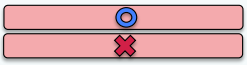
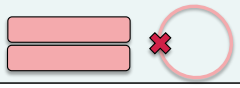

No variants on supramolcular models

| | Interface | | Surface |
|---|---|---|---|
| | Known (template) | Mined interface | |
| Polymorphism | 261 | 37 | 1,262 |
| Disease-related | 287(21%) | 44(3%) | 1,014(76%) |
| Unclassified | 218 | 22 | 714 |
| Total | 766 | 103 | 2,990 |

$p = 4.69 \times 10^{-9}$

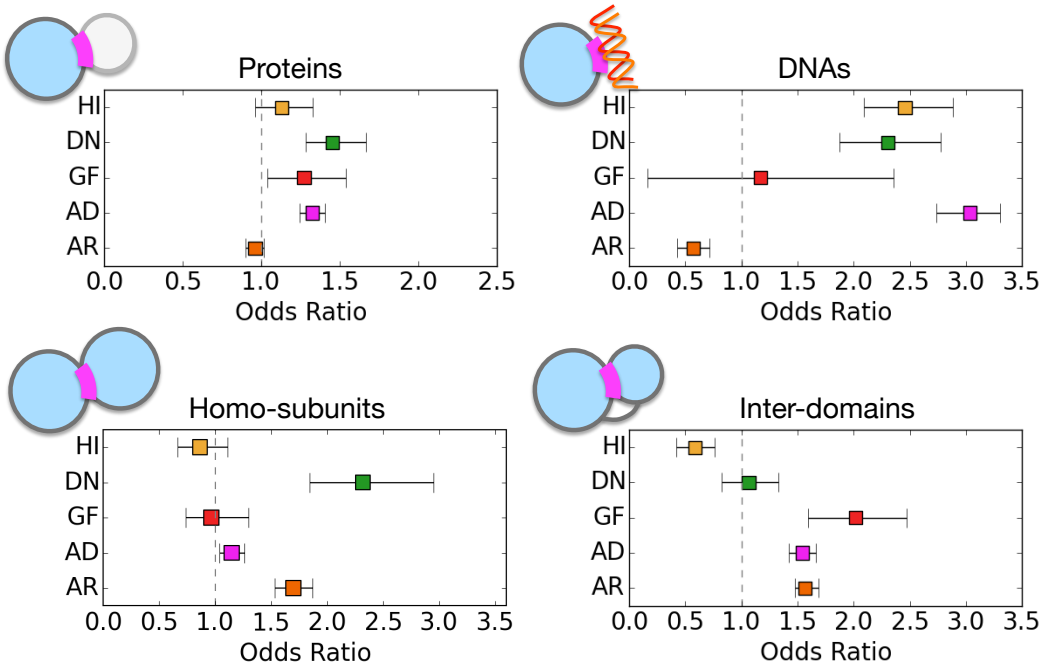
# 遺伝形式と発現形式

1) 疾患関連変異遺伝子は劣性(潜性)遺伝する場合が多いが、優性(顕性)遺伝する重大な疾患データも蓄積してきている。
2) 優性(顕性)遺伝する疾患関連変異の発現形式は比較的複雑であり、疾患メカニズムの解明が進んでいない。

| 遺伝形式 | | 発現形式 | |
|---|---|---|---|
| Recessive<br>劣性 | AR (Autosomal Recessive)<br>常染色体劣性 | |  |
| | XR (X-linked Recessive)<br>伴性(X 連鎖性)劣性 | | |
| Dominant<br>優性 | AD (Autosomal Dominant)<br>常染色体優性 | DN (Dominant Negative)<br>優性阻害 |  |
| | | HI (HaploInsufficiency)<br>ハプロ不全 |  |
| | | GF (Gain-of- Function)<br>機能獲得 |  |
| | XD (X-linked dominant)<br>伴性(X 連鎖性)優性 | |  |
| 非メンデル性遺伝 | | ミトコンドリア遺伝<br>(母性遺伝) |  |

## 疾患関連変異の発現形式と分子間相互作用の相関

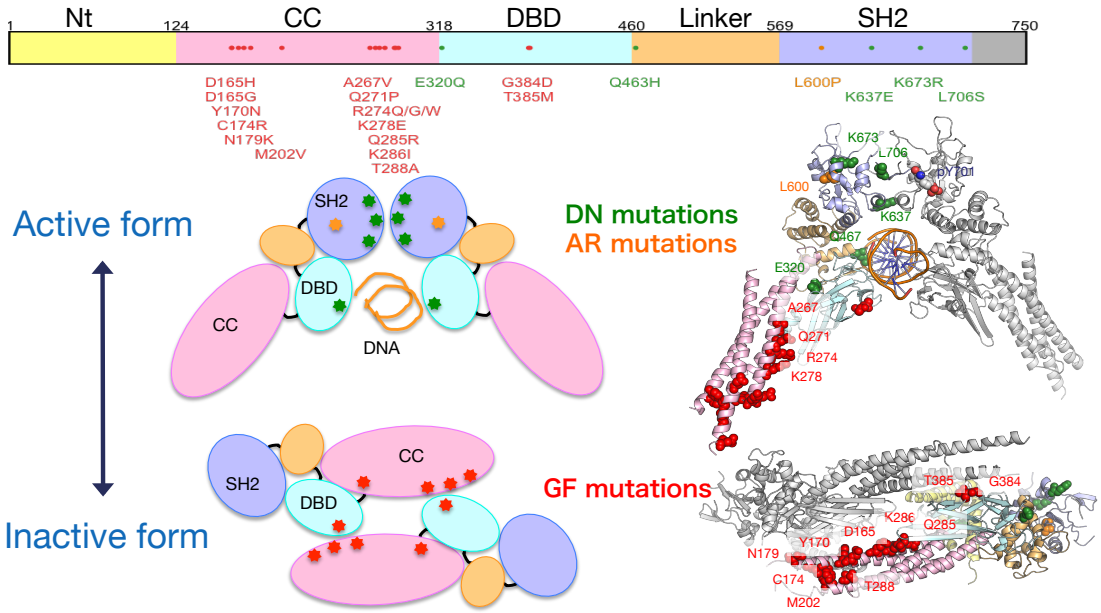
1) 発現形式とインタフェースの種類はよく相関する → 発現形式の予測に応用可能。



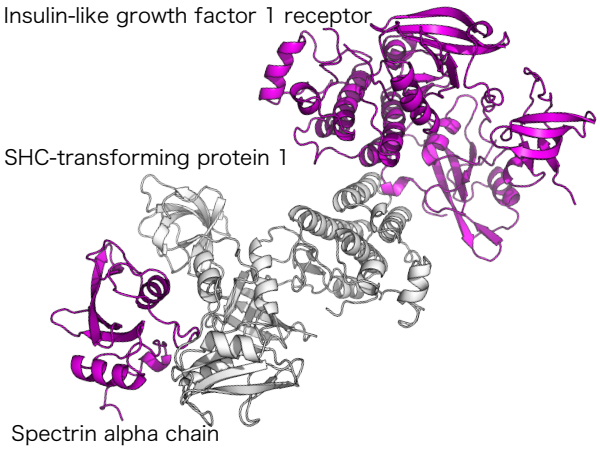

Hijikata *et al.*, *Scientific Reports* 7:8541 (2017)

# 疾患関連変異の発現形式と分子間相互作用の相関: STAT1

1) カンジダ感染症責任遺伝子(産物)STAT1上でAR，DN，GF変異はインターフェースの種類と高い相関を示す。




# 疾患関連超分子へのGWASデータのマッピング

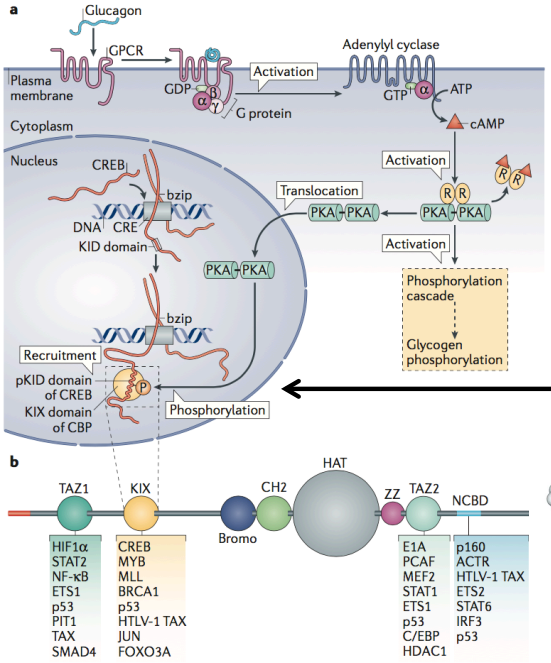



1) ～2,800のヒト超分子モデルに様々なGWAS研究の結果がマッピング可能であり、そのうち～1,200モデルは統計的有意($p < 0.01$)に特定の疾患に結びつけられる(= 疾患関連超分子)。
2) しかし、2型糖尿病の例(左図)のように、<u>複合体中心付近のサブユニットに同じ疾患関連変異がマップされない場合が多数認められる</u>。

| Gene | Detail | Essential | OMIM |
|---|---|---|---|
| NCK1 | Cytoplasmic protein NCK1 | TRUE | |
| PLCG1 | 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 | TRUE | |
| PIK3R1 | Phosphatidylinositol 3-kinase regulatory subunit alpha | TRUE | SHORT syndrome |
| ERBB3 | Receptor tyrosine-protein kinase erbB-3 | TRUE | |
| | ... | | |

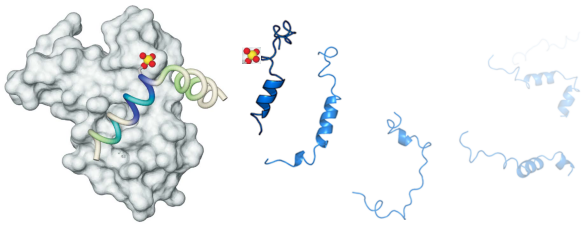
Tsuji *et al., to be submitted* (2017)

# 構造インフォマティクスの今後の課題と可能性
# IDP(天然変性タンパク質)




Wright and Dyson, *Nat Rev Mol Cell Biol.* **16**, 18–29 (2015)

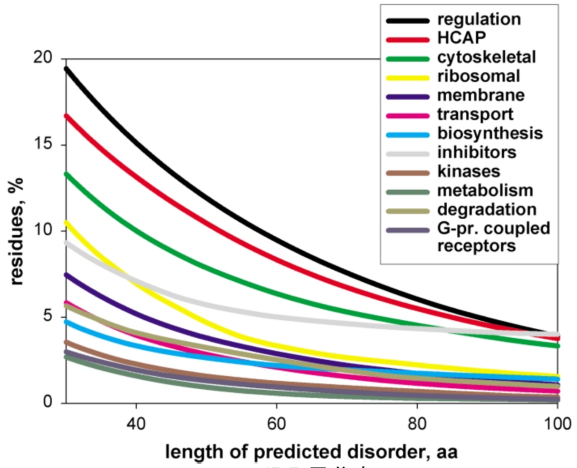
1) 天然変性タンパク質（intrinsically disordered protein；IDP）は生理的条件下で決まった三次元構造をとることができないタンパク質である。Disorder状態の「構造」は原子座標として表現が困難である。
2) 転写や翻訳といった細胞過程では重要な役割を果たしていると考えられる。このようなタンパク質は、リン酸化などの翻訳後修飾によって部分的に標的タンパク質と相互作用することが多い。
3) 多様なタンパク質と結合するためにコンフォーメーション調節できる、反応半径(リーチ)を伸ばす、標的タンパク質の入り組んだ構造にアクセス可能になるなどの構造的利点があると考えられる。



Sugase, Dyson, Wright, *Nature*, 447, 1021(2007) *modified*

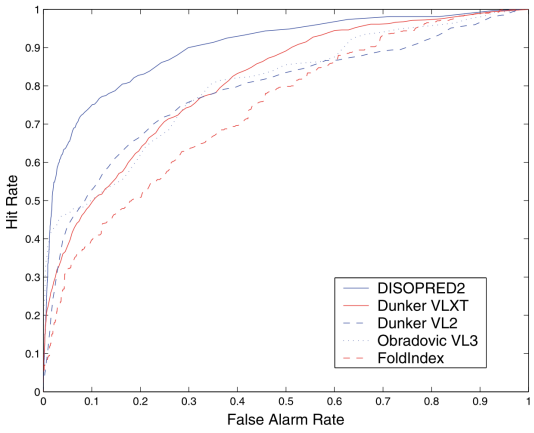
# 構造インフォマティクスの今後の課題と可能性
# IDP(天然変性タンパク質)

1) IDPは転写などの調節(regulation)など human cancer-associated proteins (HCAP)などに特に頻繁に見られ(5%程度のタンパク質が100アミノ酸を超えるIDR(intrinsically disordered region)を持ち、ゲノム/プロテオームの無視できない領域に相当する。
2) IDP/IDRの予測は、主にアミノ酸組成に基づいて行われる(疎水性残基の含有量が少なく、極性残基・荷電残基が多く含まれる傾向がある)。標準的な予測法としてはDISOPRED2 (http://bioinf.cs.ucl.ac.uk/psipred/)などがある。
3) IDP/IDRの「構造」は原子座標で表現できない。したがって現在の構造インフォマティクスに融合しにくい。




IDR長分布

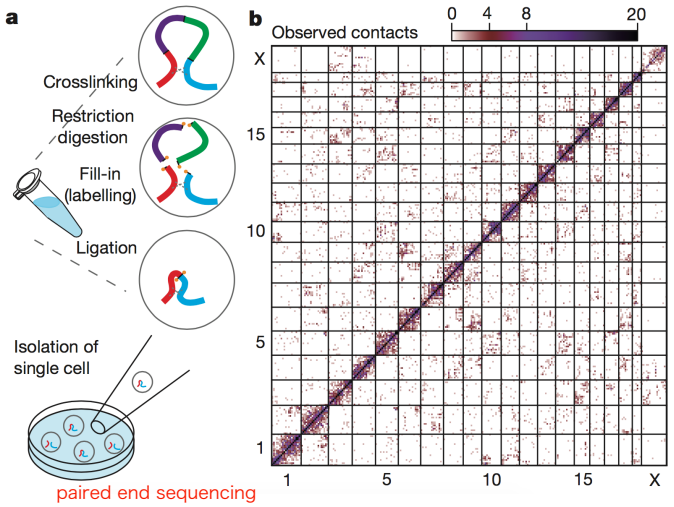
Lakoucheva *et al.*, JMB, 323, 573 (2002)




IDR予測精度

Ward *et al.*, JMB, 337, 635 (2004)

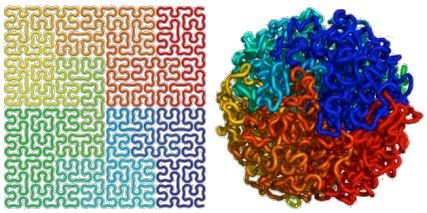
# 構造インフォマティクスの今後の課題と可能性
## Hi-C(染色体立体構造)




Hi-C (all vs all HTS Chromosome Conformation Capture )

Nagano *et al*., *Nature,* **502**, 59 (2013)

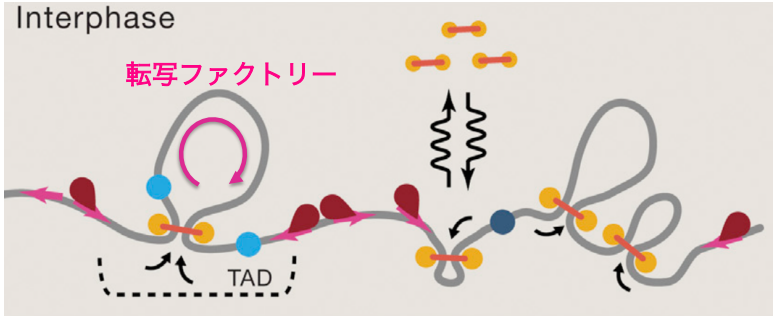
1) 3C法（chromosome conformation capture）を基盤にしたHi-C法は空間的に接近したゲノムDNA(染色体テリトリー)をライゲートしpaired-end配列解析と距離地図により染色体の「立体構造」を解明する手法である。
2) 単一細胞Hi-C法では単細胞で解析を行い、X染色体三次元構造モデルの構築に成功している。染色体は一定の立体構造をとるわけではないが、高頻度で観測される染色体テリトリー・ドメイン構造や、染色体間の相互作用が存在する。
3) 染色体はフラクタル構造(ヒルベルト曲線)を取るとされる。

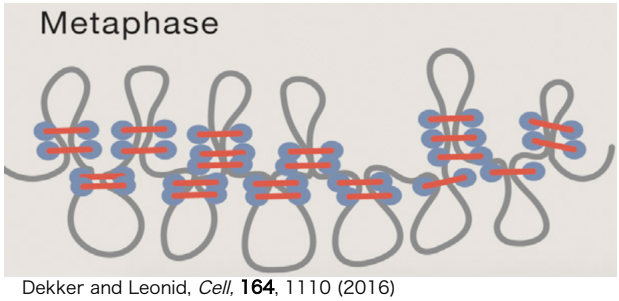


Nagano *et al*., *Nature,* **502**, 59 (2013)

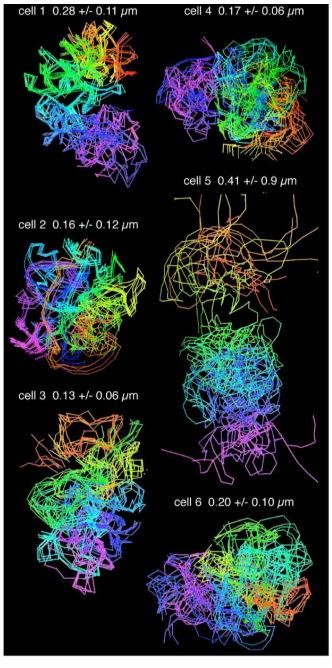
# 構造インフォマティクスの今後の課題と可能性
## Hi-C(染色体立体構造)

1) 染色体立体構造は転写ファクトリーの形成などにも関与するが、これらのデータもまだ構造インフォティクスにとり込めてはいない。




Dekker and Leonid, *Cell,* **164**, 1110 (2016)




X染色体"コンフォメーション"

Nagano *et al*., *Nature,* **502**, 59 (2013)