

ポストゲノム時代の計算生命科学

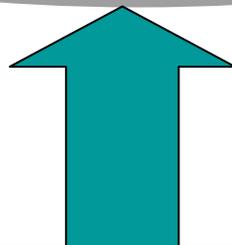
郷 信広

原研計算科学技術推進センター量子生命情報解析グループ

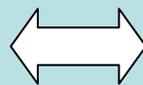
中性子利用研究センター生体物質機能解析研究グループ

奈良先端大バイオインフォマティクス人材養成ユニット

ポストゲノムの生物学



バイオ
インフォマティクス



大規模生体高分子系
シミュレーション

ゲノム

遺伝子 → 要素機能

その発現制御

たんぱく質、RNA

転写制御因子

機能統合

相互作用

高次機能

高次構造体

情報に基づく生物学

実体に基づく生物学

バイオインフォマティクス

物理学、化学

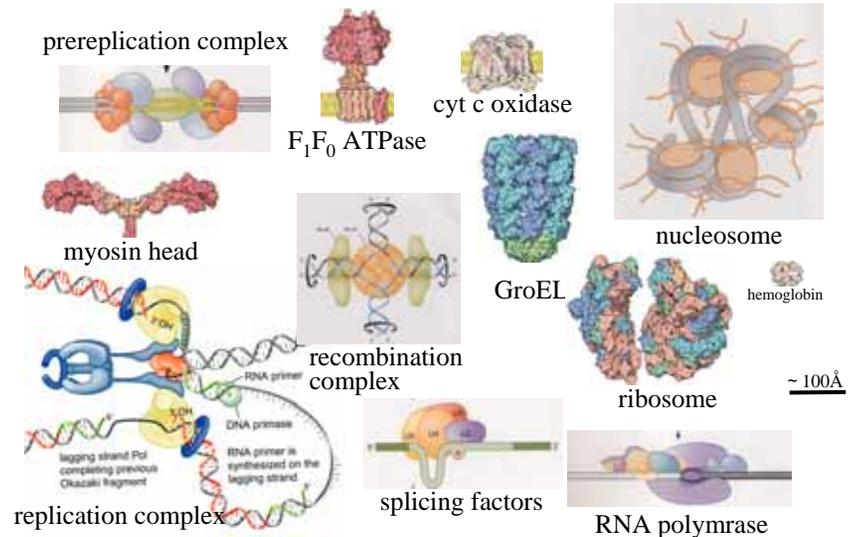
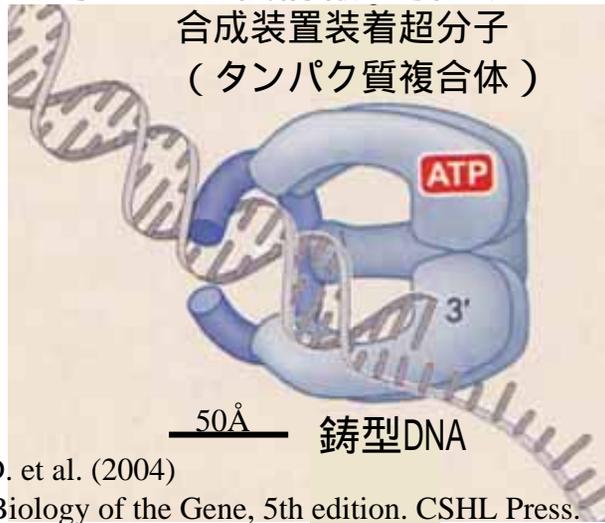
ゲノム決定後の生命科学研究

システムバイオロジー

生体高分子の相互作用にもとづく生命現象

1. タンパク質が相互作用して超分子を形成
2. タンパク質の弱い相互作用による情報伝達ネットワークの形成

例) 親から子への遺伝情報の引き継ぎシステム



1. 生命を化学のことばで理解・説明
2. タンパク質の異常による病気の原因を原子レベルで解明
3. 生体超分子の機能実現方法を理解し、材料設計に応用

バイオインフォマティクス

基本論理 = 進化 (物理学・化学を補完)

弱い経験則の集まり

経験則の一般形

「あるデータベースDを構成する要素のなかでFと言う(立体構造上の、あるいは機能上の)性質を持つものの中には、Pと言う(配列上の、立体構造上の、あるいは物性上の)特徴を持つ一群のものがある」

必要条件でも十分条件でもない → 弱い法則

法則の弱さの原因

1. Fという性質を持つものに、いくつかのタイプがある。(異なる進化起源)
2. 個々のタイプの境界が明確にし難い。

弱さの克服

必要性の向上: Fという性質を持つものの中に、新しい特徴P2を持つものを探す。

十分性の向上: 複数の法則を同時に満たすものを検索。

統合バイオインフォマティクスシステム

Bioinformatics:

Ask

Any

Questions

経験則一つ一つを検索シナリオとして蓄積

(原研計算科学: 由良敬、奈良先端大: X. Gong, J. Cheema, 中村健介)

バイオフィンフォーマティクスの直面する状況

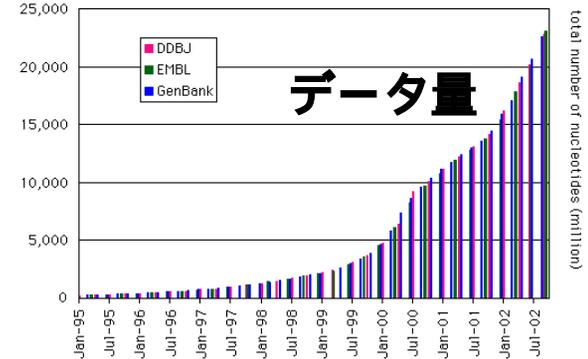
- 1 . データの爆発的な増加
- 2 . 手法とデータフォーマットの増大
- 3 . コンピュータ資源の制約



どの方法を使ってどのように解析すれば、
疑問に回答を得ることができるのかわ
からない。

バイオフィンフォーマティクスにおける疑問に対し
て、簡単に回答をえたい。

DDBJ/EMBL/GenBank database growth 1995/1-2002/9



BAAQ システム: 統合バイオインフォマティクスシステム (Bioinformatics: Ask Any Questions)

これからのバイオインフォマティクスシステムには以下の機能が必要;

1. 解析手段が概観できるシステム、
2. データのフォーマットを気にしなくてもよい状況の実現、
3. 解析方法の使用方法が簡単にわかる状況の実現 (知識の集約)
4. ユーザが意識せずに異なる計算機資源を利用できる環境の実現

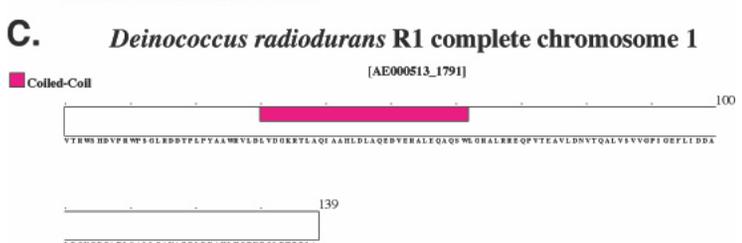
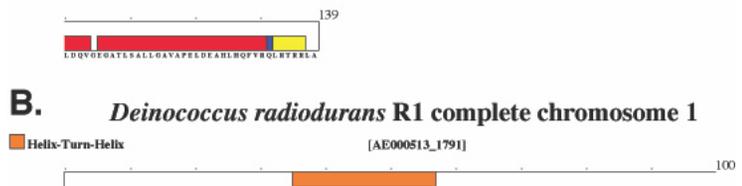
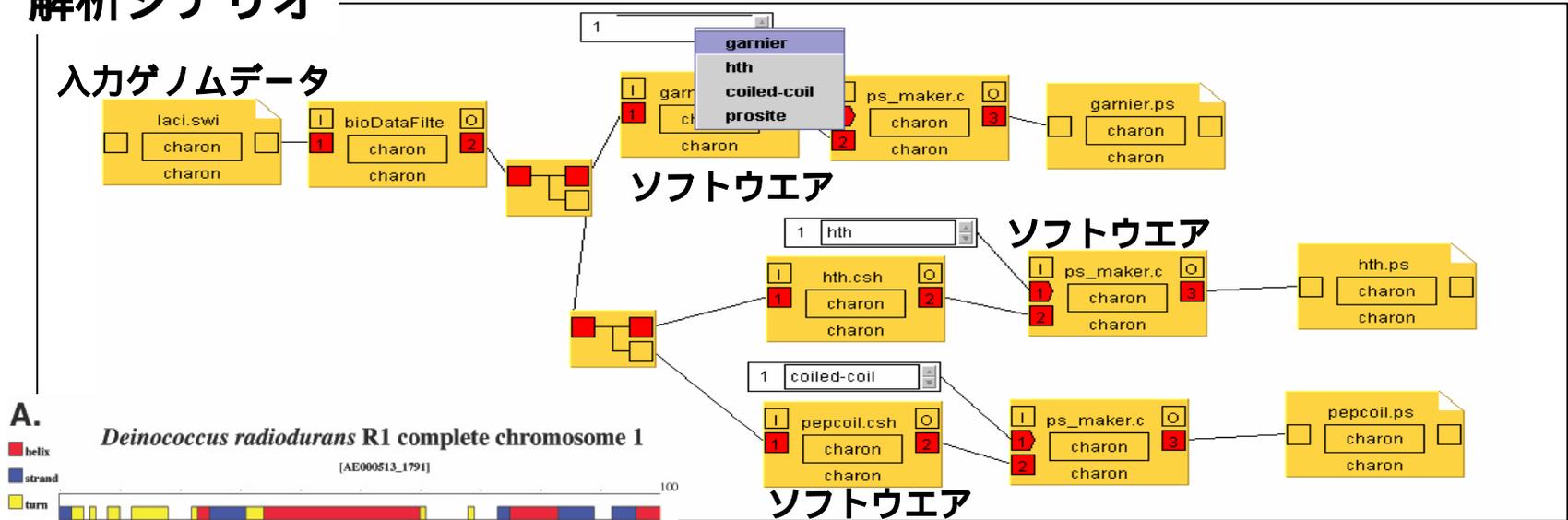
The image displays a screenshot of the BAAQ system interface. On the left, a window titled 'gv: phylip.ps' shows a phylogenetic tree with the title '結果' (Result). The tree has a root node that branches into two main groups. The left group contains nodes 'gi 515461' and 'gi 5737802'. The right group contains nodes 'gi 5739002', 'gi 671595', 'gi 1722465', and 'gi 3255938'. On the right side, a 'Task Management Editor (TME)' window is open, titled 'プログラムとデータ' (Program and Data). It shows a hierarchical task graph with various colored boxes representing tasks and data. Tasks include 'in_clw', 'sch2clw', 'clustalw', 'neighbor', 'drawtree', 'gv', 'in_cos', 'in_dtr', 'neighbor', 'consense', and 'vitermsh'. Each task box contains a small thumbnail image and some numerical values. Arrows indicate dependencies between tasks. A terminal window at the bottom right shows the IP address '172.16.165'.

原研で開発しているSTA (Seamless Thinking Aid)上で、上記の基盤を構築する

例えば、

ゲノムからDNAと相互作用する遺伝子を見いだす

解析シナリオ



3個の解析用プログラムと表示用プログラムを同時に用いる必要がある。

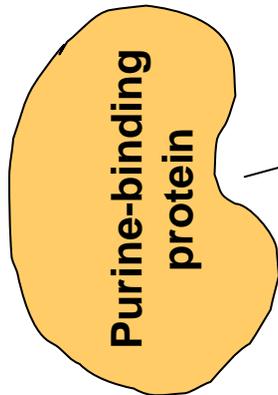
↓

解析シナリオを構築または、すでに存在するシナリオを改変利用。

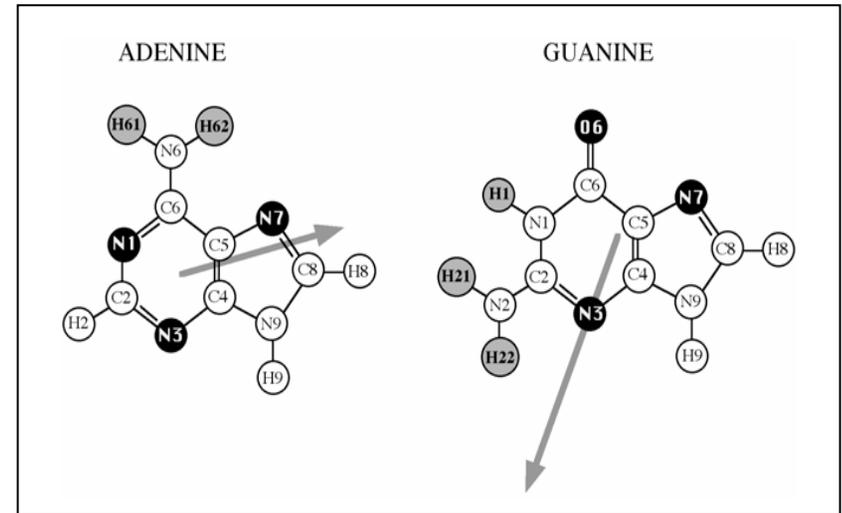
↓

データを入力して解析完了。

Electrostatic potentials and Biological Specificities



BINDING SITE:
Adenine-specific?
Guanine-specific?
Non-specific?

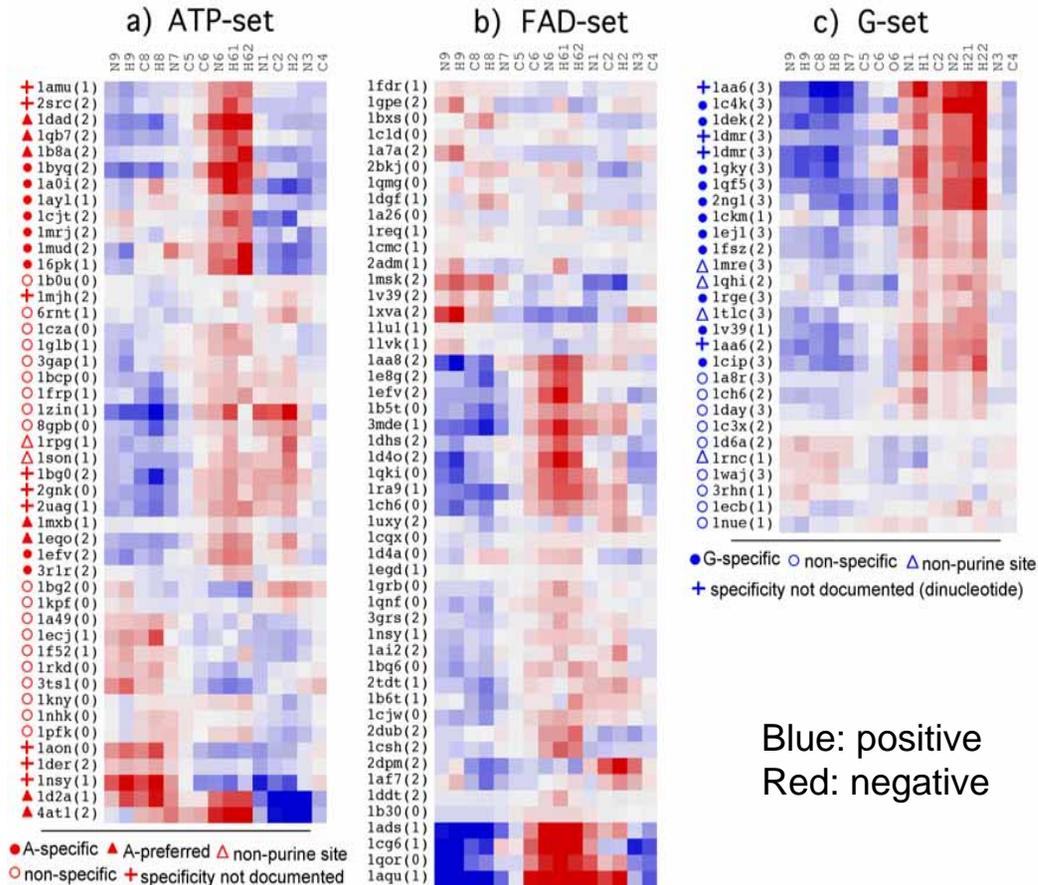


Electrostatic potential of nucleotide-free protein is sufficient for discrimination between adenine- and guanine-binding sites (*J. Mol. Biol.* 2004, 342:1053-1066)

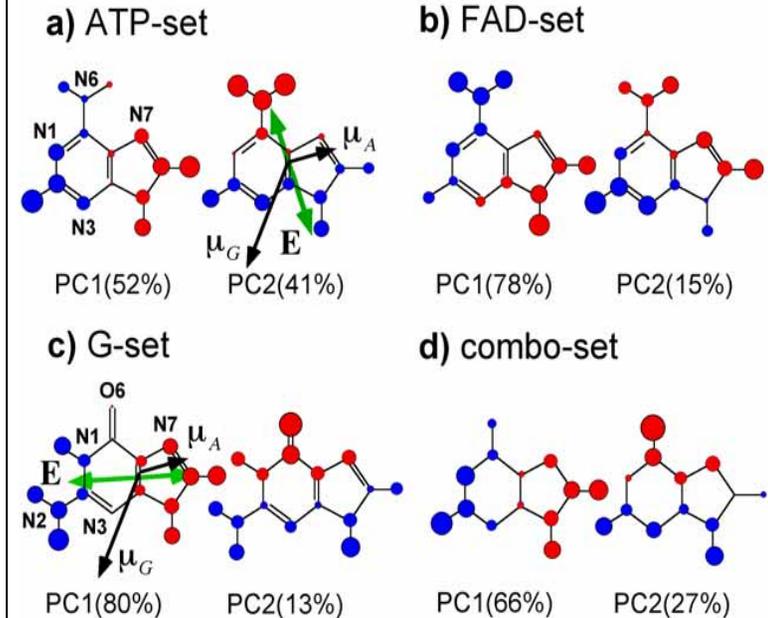
(奈良先端大: G. Basu, D. Sivanesan, 川端猛)

Analysis of electrostatic potentials

Electrostatic potential (zero-mean) at purine-binding sites of 125 non-homologous proteins

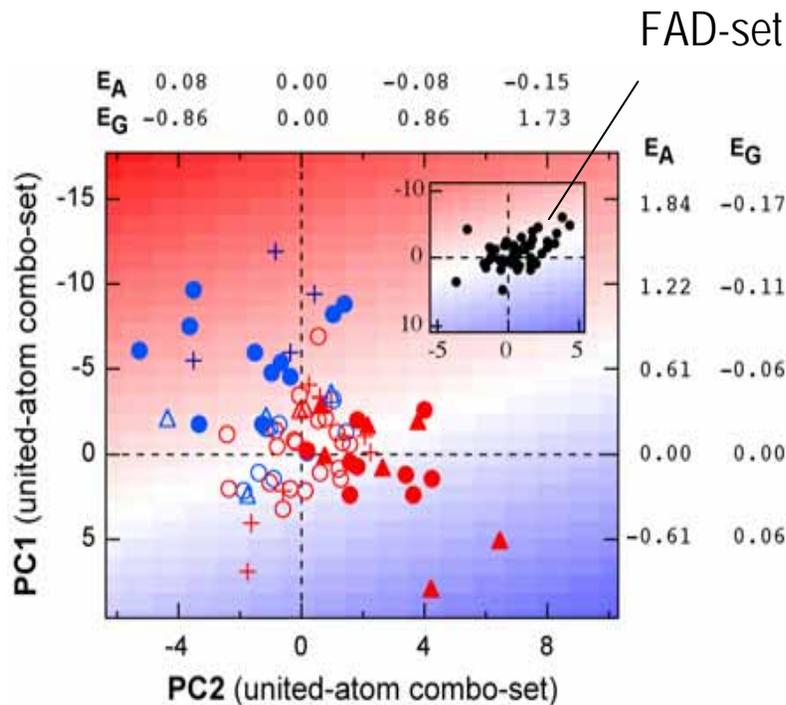


Principal component vectors

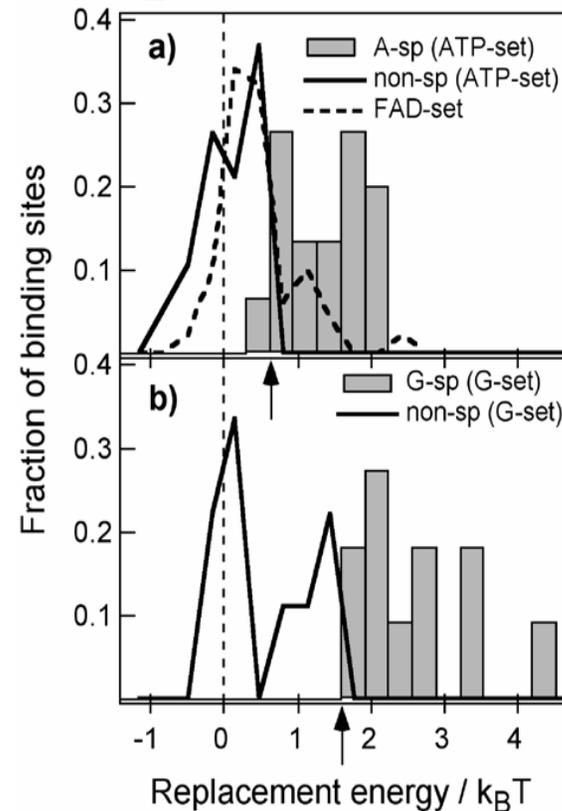


Separation of binding sites: projection on PC vectors and replacement energies

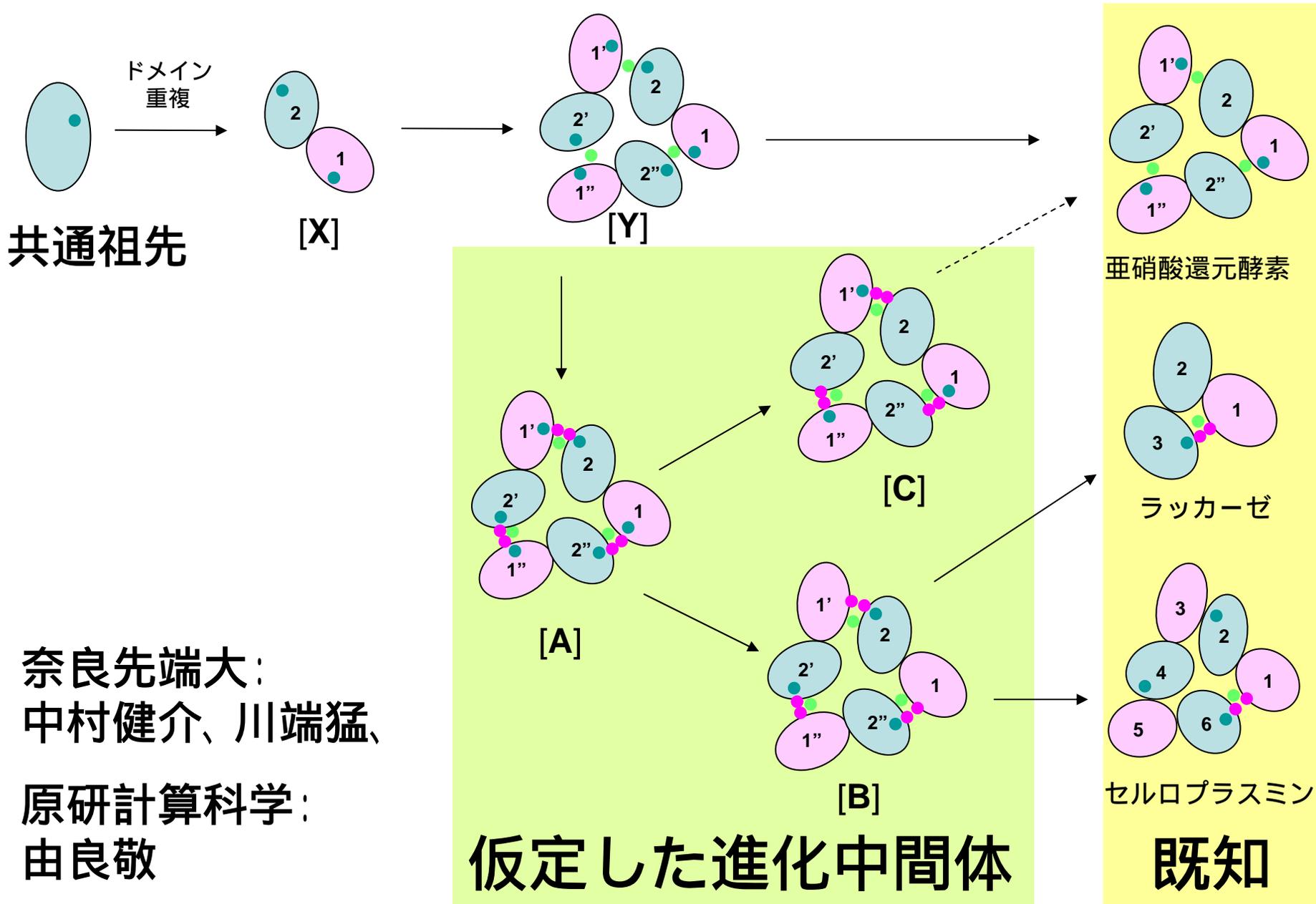
Projection of ESP of binding sites on the PC1-PC2 plane



Replacement ($A \leftrightarrow G$) Energies



マルチ銅オキシダーゼの進化過程の推定



奈良先端大：
中村健介、川端猛、

原研計算科学：
由良敬

進化モデルで仮定した中間体をゲノムデータベース中に発見！

II

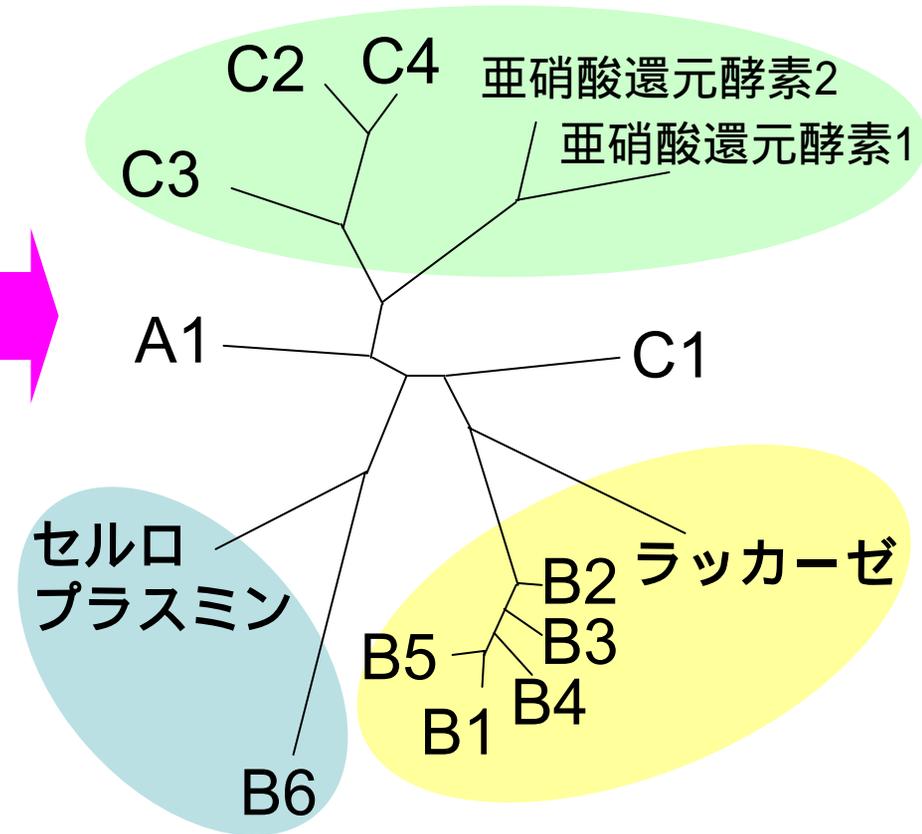
新しいタイプの銅結合蛋白質

亜硝酸還元酵素
ラッカーゼ
セルロプラスミン

既知

- A1 : 藍藻
- B1 : 根粒菌
B2 : アニリン資化性菌
B3 : 根粒菌
B4 : 磁性細菌
B5 : 根頭癌腫菌(薔薇)
B6 : ストレプトマイシン産生菌
- C1 : 硝化細菌
C2 : 藍藻
C3 : 家畜の病原菌
C4 : 藍藻

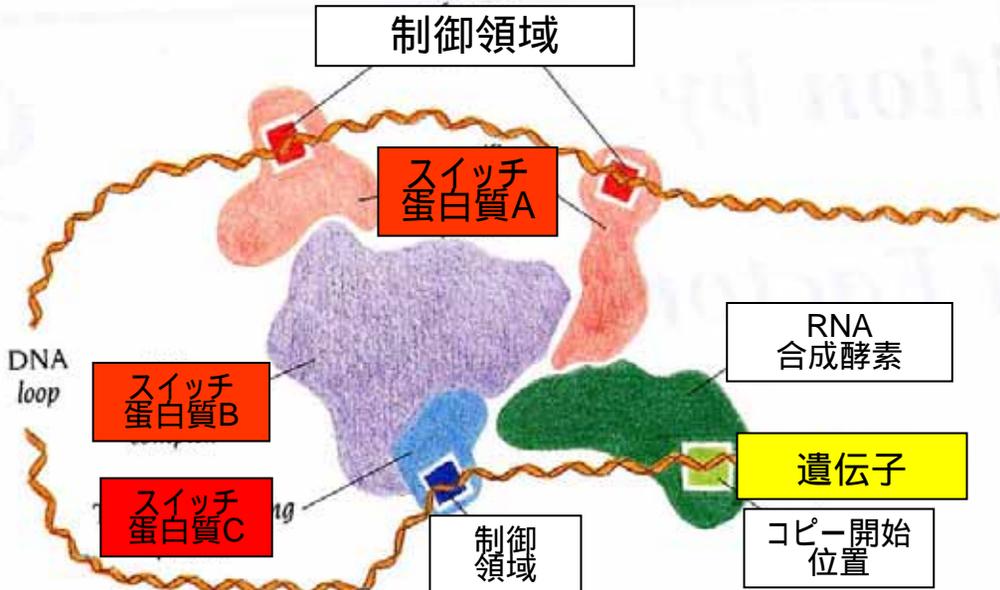
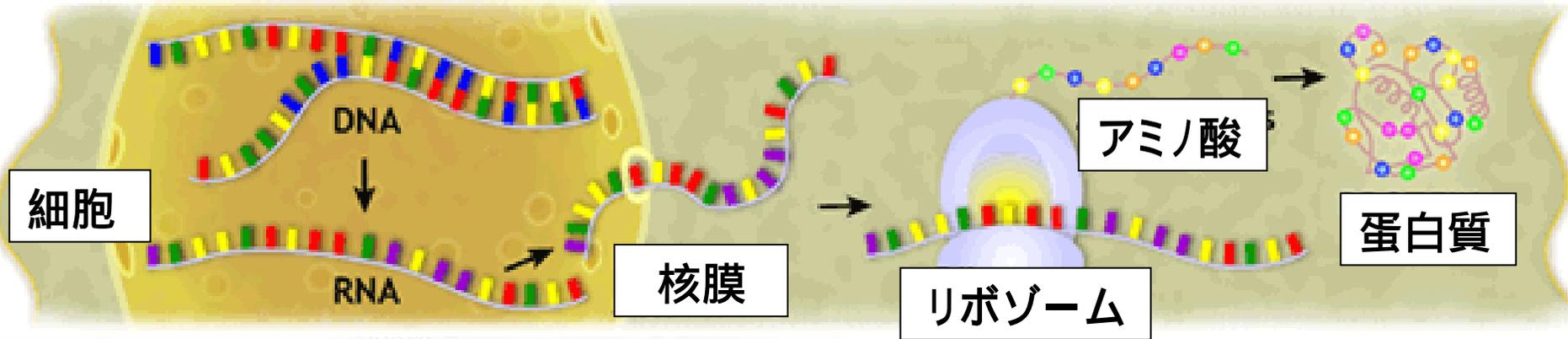
新規



発見した配列から作成した系統樹が進化モデルをサポート

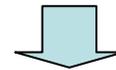
遺伝子発現から蛋白質生産まで

セントラルドグマ: DNA(遺伝子) → mRNA → 蛋白質
コピー(転写) 翻訳



スイッチ蛋白質 (転写制御因子) の役割

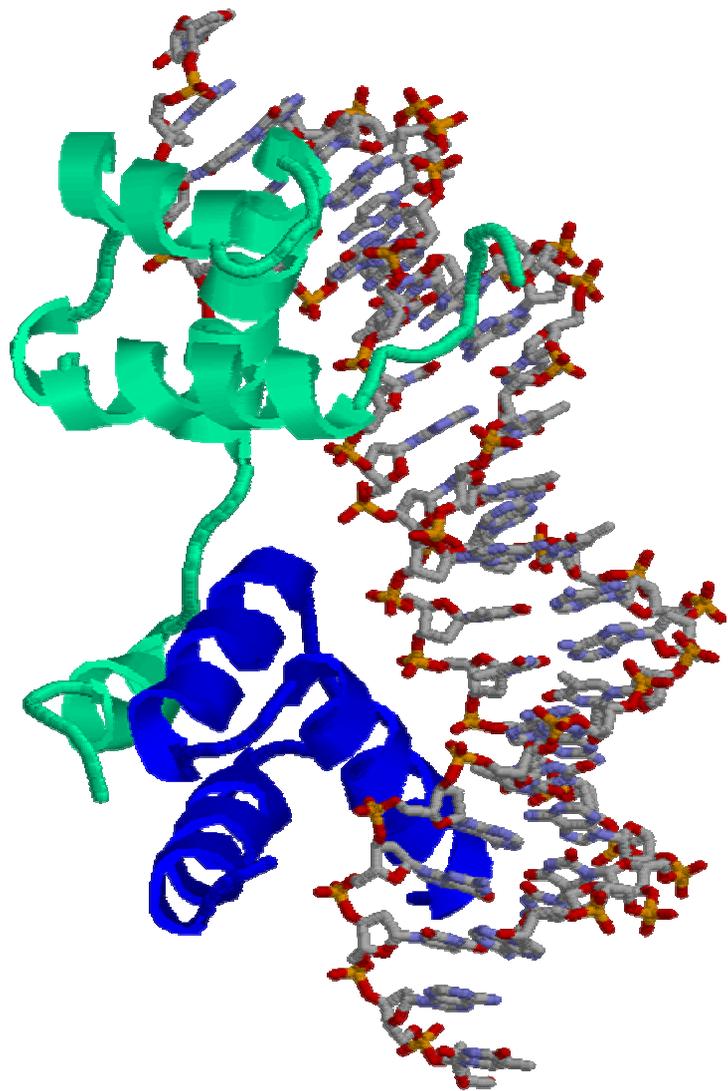
特定のDNA配列を認識し、結合



コピーのオン、オフの制御

(原研中性子: 河野秀俊)

スイッチタンパク質のDNA配列結合特異性の解析



直接認識

アミノ酸(蛋白質)と塩基(DNA)の直接的な相互作用による認識



蛋白質 - DNA相互作用の
経験的ポテンシャルの構築

間接認識

直接相互作用していない塩基置換による結合強度の変化

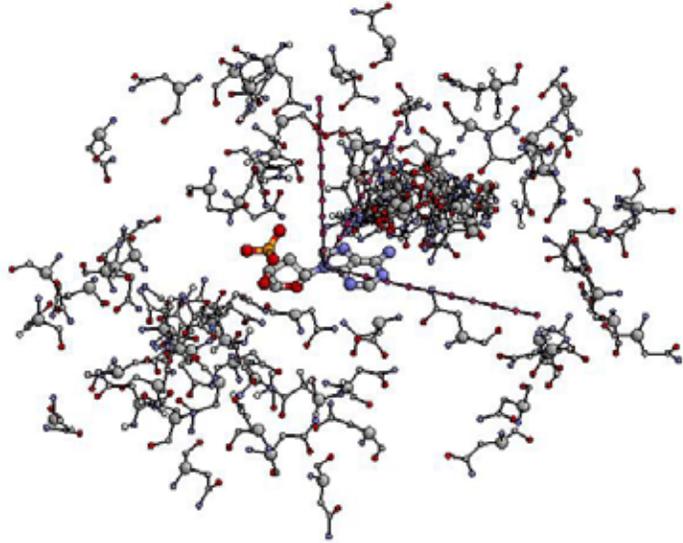
DNAの構造変形のしやすさの認識



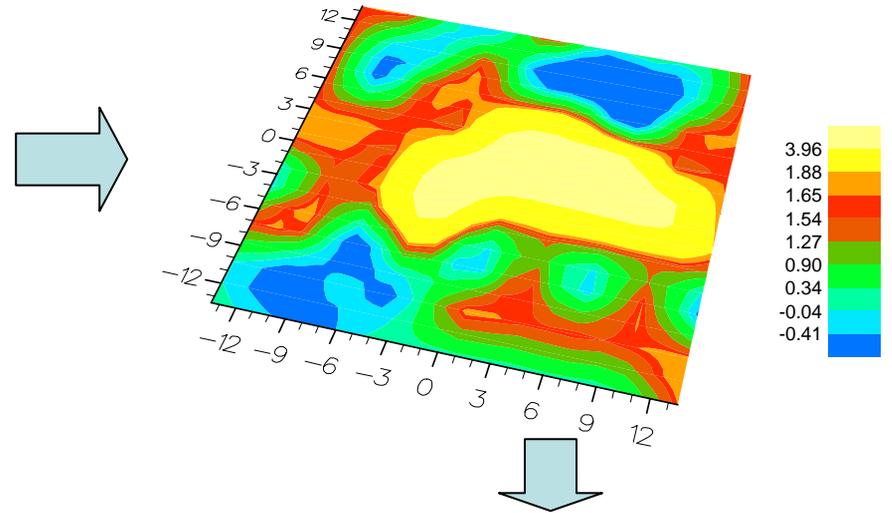
構造変形しやすさの
DNA配列依存ポテンシャルの構築

直接認識

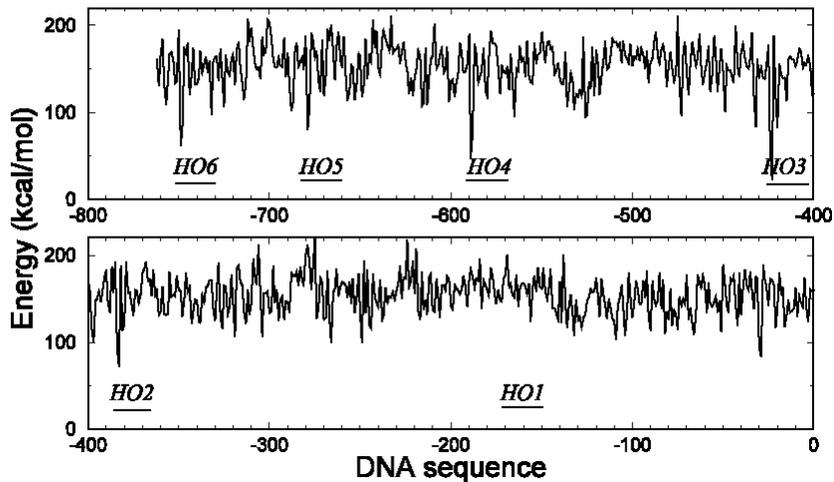
塩基 (DNA) のまわりのあるアミノ酸 (アスパラギン) の空間分布



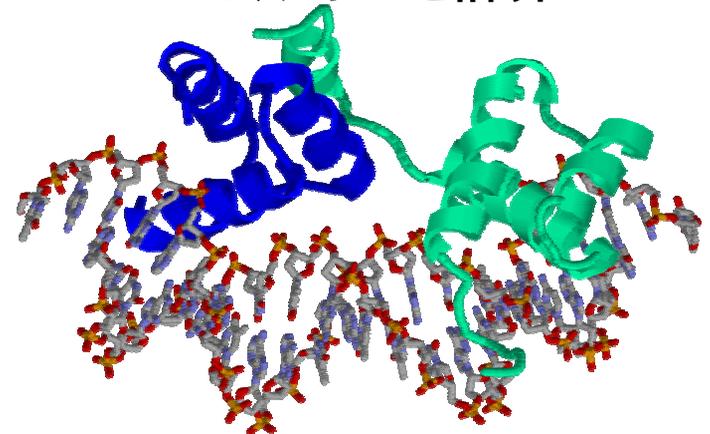
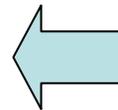
統計ポテンシャル: $\Delta E_{ij}(s) = -\ln(f_{ij}(s)/f(s))$
4塩基 \times 20アミノ酸 = 80の組み合わせ



ゲノム上での標的配列の認識

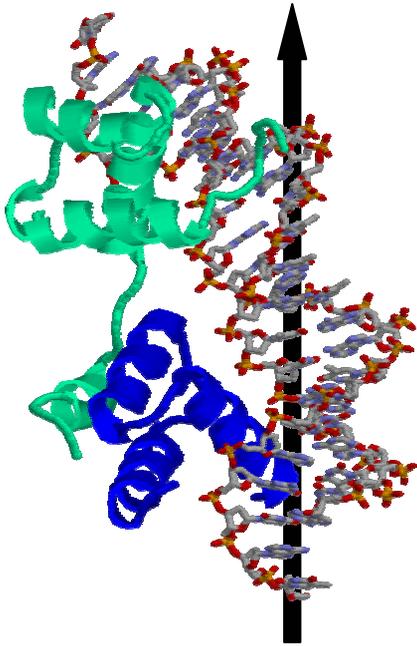


DNAの配列を載せ代えて、
エネルギーを計算

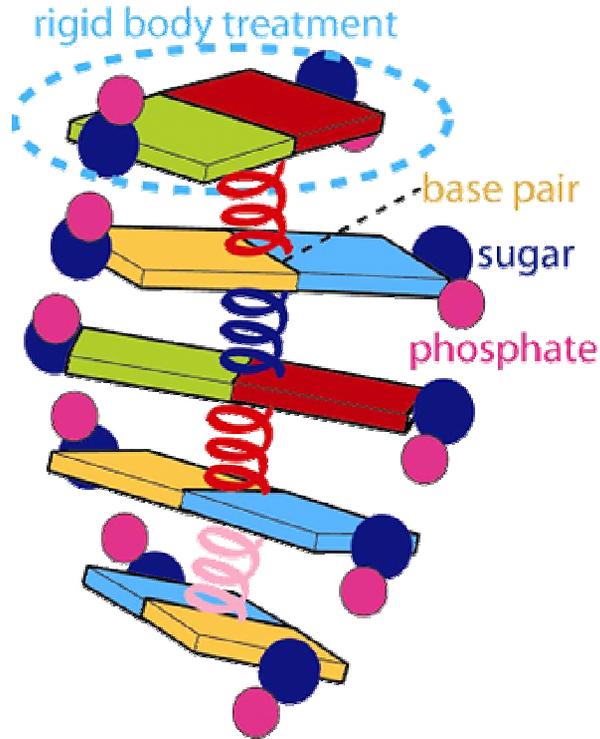


間接認識

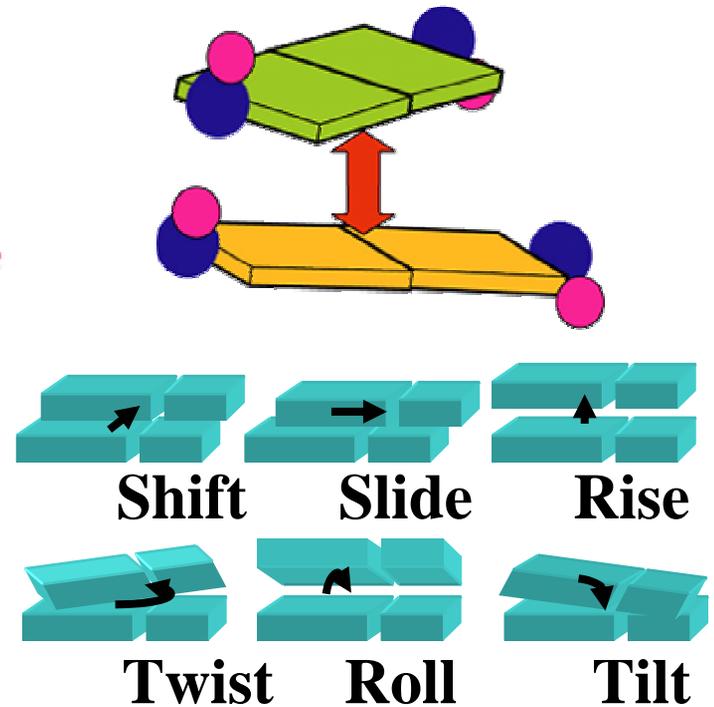
DNAの曲がり



塩基対の剛体近似



DNAの構造パラメータ (2塩基対)



逆擬調和解析 (Inverse quasi-harmonic analysis)

$$\langle \Delta \theta_i \Delta \theta_j \rangle = k_B T [\mathbf{F}^{-1}]_{ij}$$

\mathbf{F} : 力定数行列

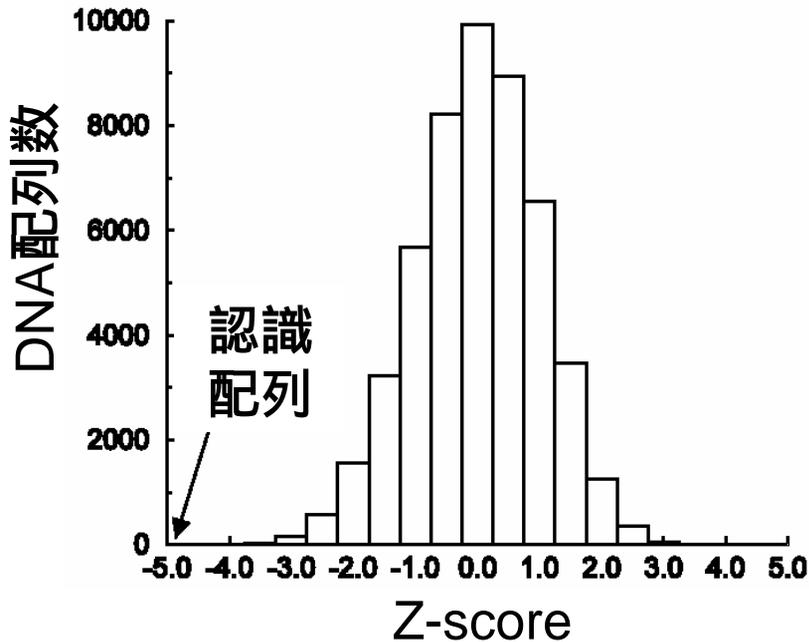
$$\langle \Delta \theta_i \Delta \theta_j \rangle = \langle \theta_i \theta_j \rangle - \langle \theta_i \rangle \langle \theta_j \rangle$$

共分散行列

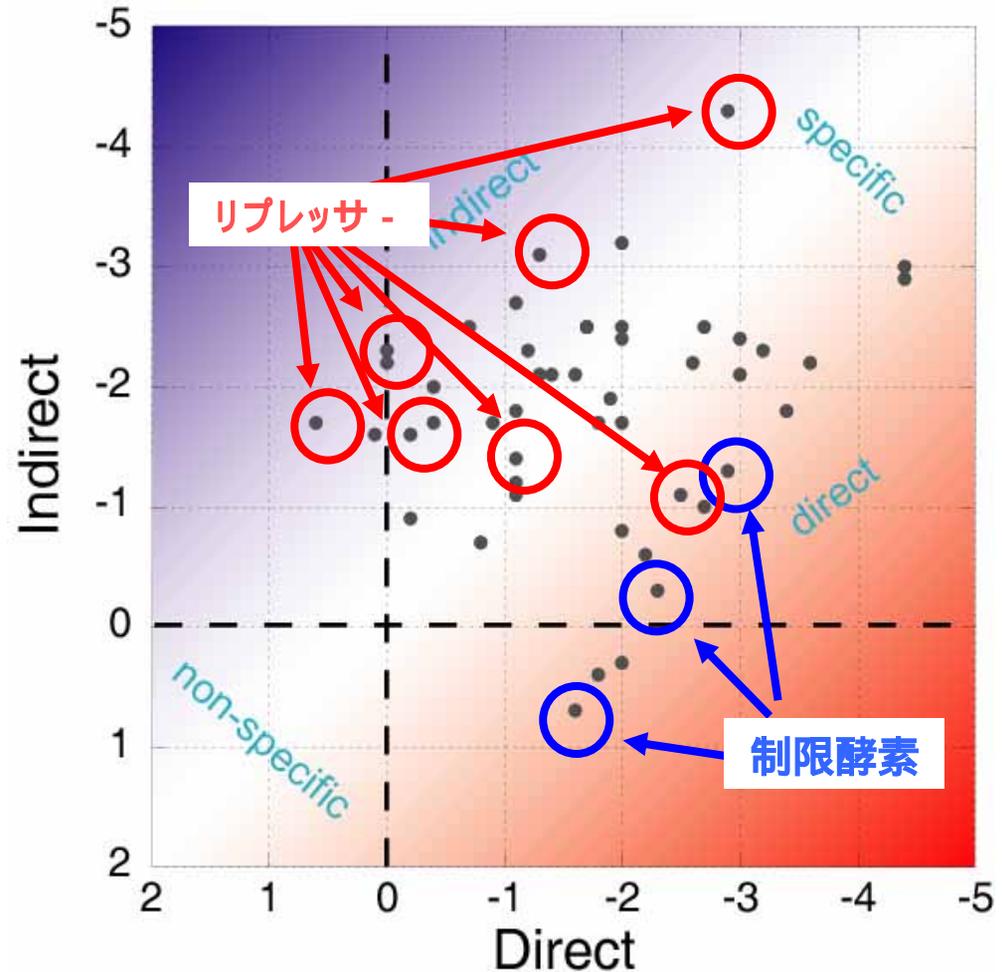
配列認識における直接認識と間接認識の寄与分解

認識配列をランダム配列とのエネルギー差として評価

$$Z\text{-score} = \frac{X - m}{\sigma}$$



52のDNA結合蛋白質の認識配列のZ-score

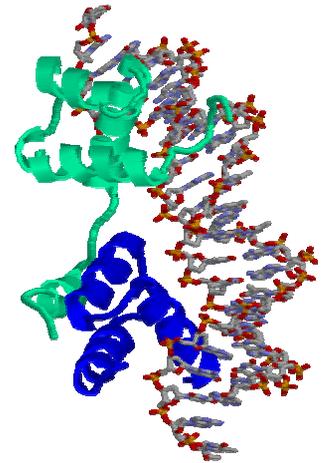


特異的なDNA配列に結合する蛋白質の設計システム開発

- スイッチ蛋白質の解析と設計 -

河野 秀俊

- 既知スイッチ蛋白質が認識するDNA配列とアミノ酸の相互作用解析
 - アミノ酸 - DNA相互作用のポテンシャル関数の導出
 - 二つの応用
 - スイッチ蛋白質の標的遺伝子の同定
 - スイッチ蛋白質の設計
- 任意のDNA配列を認識するスイッチ蛋白質の設計
 - 既知スイッチ蛋白質のDNA認識配列の改変(アミノ酸残基置換)
 - アミノ酸配列改変にかかわらず、蛋白質の安定性の維持、向上
 - 任意の9塩基対 ($4^9=262,144$ 通り) を特異的に認識



*Deinococcus radiodurans*はなぜ放射線に強いのか？

D. radiodurans は放射線に対して、ヒトよりも1,000 倍以上強い。放射線によってDNAは切断されるが、すぐに修復されることが知られている。高性能修復をつかさどる因子は？

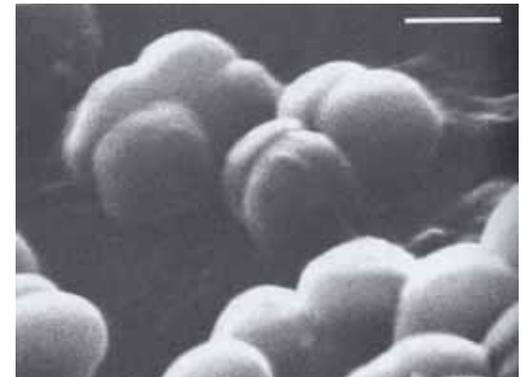


ゲノム決定プロジェクトによって、

- 1) 特別な遺伝子を発見できなかった、
- 2) 機能がわからない遺伝子がたくさんあることがわかった (1,694個)。



新規修復遺伝子の存在が期待される。
しかし、従来の機能予測方法では、
発見できない。

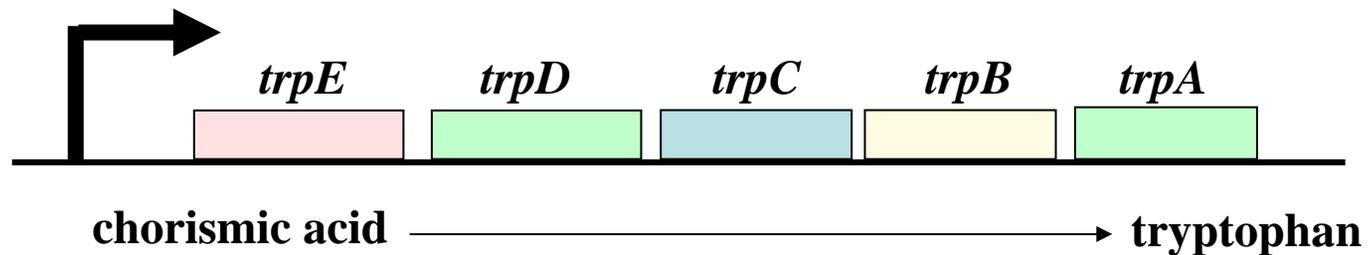


An electromicroscopy picture of *Deinococcus radiodurans*, a radiation resistant bacterium. (L. Margulis & K.V. Schwartz, "Five Kingdoms" (1997) p.100)

遺伝子クラスター（オペロン）構造にもとづく 遺伝子機能予測

オペロン：バクテリアに存在する遺伝子転写ユニット。
複数の遺伝子が同時に転写され、ひとつのmRNAになる。

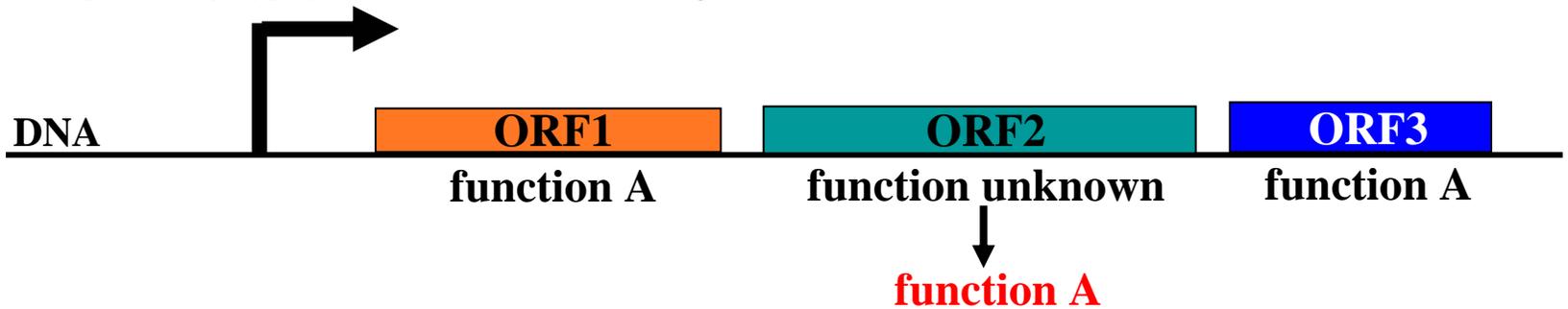
大腸菌ゲノムにみられる有名なオペロン トリプトファン オペロン



同じオペロン内の遺伝子は機能に関連がある場合が知られている。

機能予測への適用

1. オペロン構造を予測、
2. 既存の方法で遺伝子の機能予測、
3. 同じオペロン内に存在する「機能未知」遺伝子に、オペロン内の機能をコピーする。



どの程度の精度で機能を推定できるか

D. radiodurans genome (chr 1 (2579 ORFs))

all similar	23	→ 67
partially similar, partially different	8	
partially similar, partially hypothetical	36	
all different	44	
all different, partially hypothetical	199	
all hypothetical	148	

60% 程度は正しい。

バイオインフォマティクスによる予測と実証実験

✗	116 Uracil-DNA glycosylase	
	116 DNA topoisomerase I	
	116 Hypothetical protein	cytoplasmic (254aa)
	84 DNA polymerase III alpha subunit	
	84 Hypothetical protein	cytoplasmic (170aa)
	382 DNA polymerase III, delta' subunit	
	382 Hypothetical protein	cytoplasmic (257aa)
●	182 DNA repair protein recF	
	182 Hypothetical protein	cytoplasmic (288aa)
●	36 Recombination protein recR	
	36 Hypothetical protein	cytoplasmic (203aa)
✗	71 endodeoxyribonuclease ruvC	
	71 Hypothetical protein	TM2-4 (166aa)
	⋮	

線照射前 | 線照射後

野生株  
耐性

遺伝子破壊株  
感受性

原研高崎研鳴海博士らによる実験

2個の機能未知遺伝子が放射線抵抗性細菌のDNA修復に関与していることを明らかにした。

巨大生体高分子のシミュレーションシステム PABIOS (PARALLEL BIOMOLECULAR SIMULATOR)の開発

[シミュレーションの現状]
生体高分子系 (1~2分子)
10⁴粒子、10⁶ステップ
のシミュレーション

スーパーコンピュータ

大規模化、高速化

生体高分子複合体系
10⁶粒子、10⁶ステップ以上
のシミュレーション

PABIOSの高速化・高精度化

並列化 多数(100以上)のCPUを使用
空間分割法による通信コスト削減

動的ロードバランス
各CPUのロードバランスを動的に最適化

長距離相互作用 計算量の多い部分を高速化
PPPM (Particle-Particle Particle-Mesh)法の採用

高精度の時間積分法を採用
Multi Time Step法による時間ステップ幅の増大

PABIOSの計算性能

高い並列化効率及びベクトル化率を達成

計算負荷が高い長距離静電相互作用計算ルーチンを重点的に並列ベクトルチューニング

地球シミュレータにおける計算性能

160個のCPUによる計算

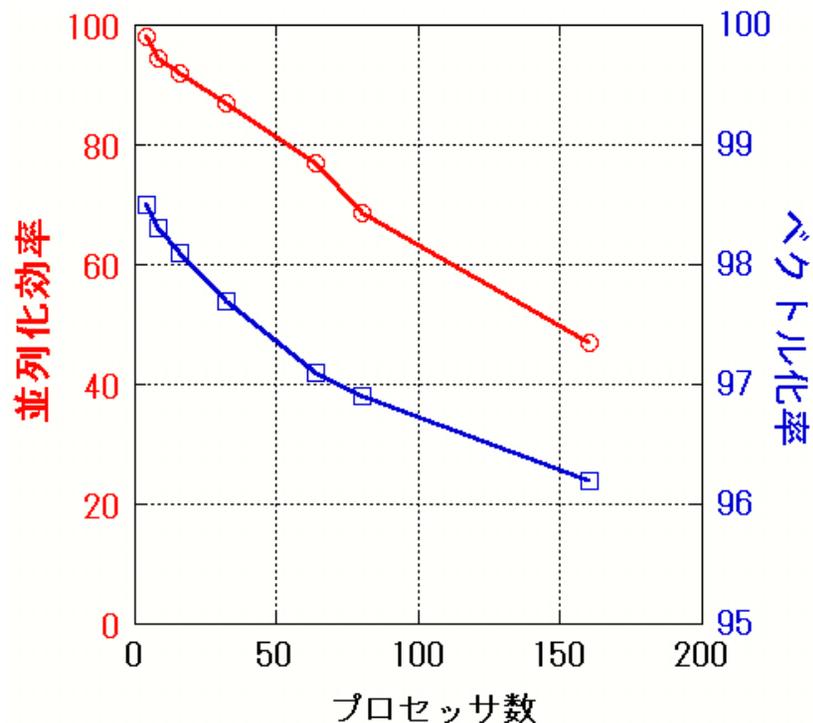
並列化効率47.6%

ベクトル化率96.2%

原研中性子: 石田恒、樋口真理子

原研計算科学: 松本淳、米谷佳晃

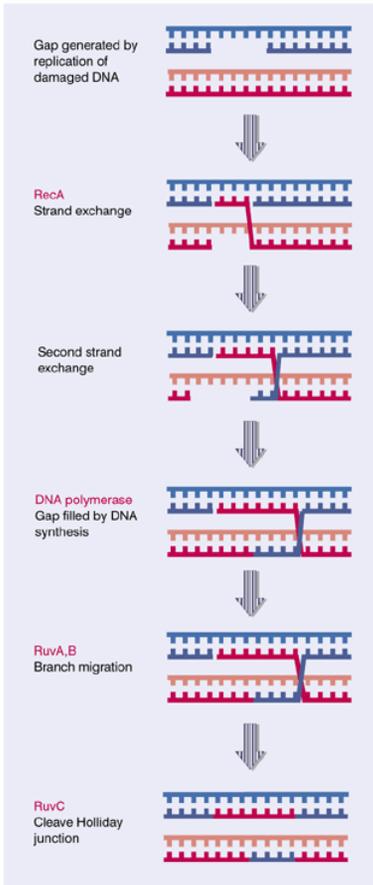
東大文生研: 北尾彰朗、城地保昌



蛋白質 + 水の系(全原子数:166,177)におけるベンチマークテスト

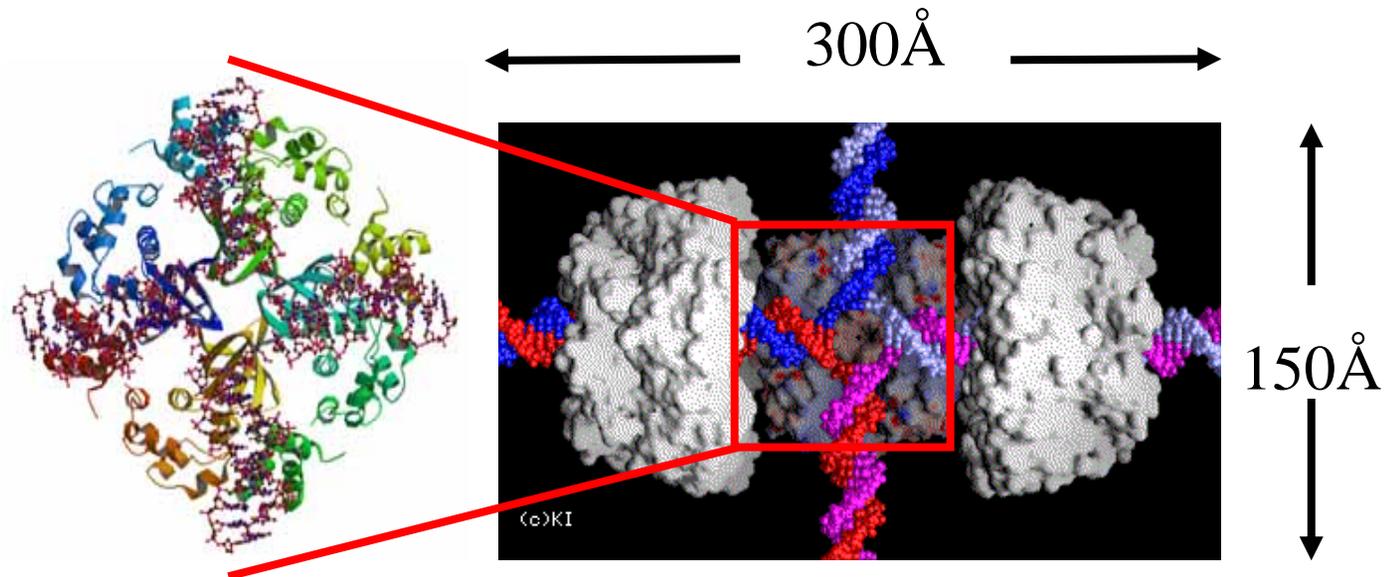
PABIOSを用いた大規模生体超分子系の機能解析

Figure 14.13 Bacterial enzymes can catalyze all stages of recombination in the repair pathway following the production of suitable substrate DNA molecules.



DNA組換えのダイナミクス解明

DNA組換えのモデル



DNAに結合したRuvAB蛋白質

DNAに結合したRuvA4量体蛋白質の系

(RuvA: 12,628 原子, DNA: 3196 原子, 水: 32253 × 3 原子, Mgイオン: 54 原子の合計11,2637原子)

West, SC. Cell, (1998)

DNAの組換え

(原研中性子: 石田恒)

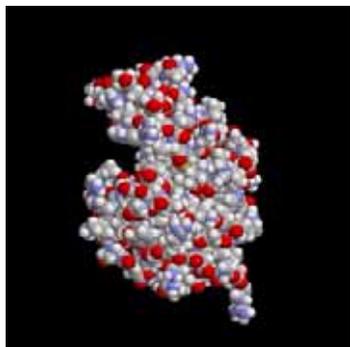
蛋白質1分子から超分子の機能解明へ

分子シミュレーションによる生体分子系の機能研究

5年前の系

酵素蛋白質
リゾチーム1分子

分子量14,000
全系2万原子

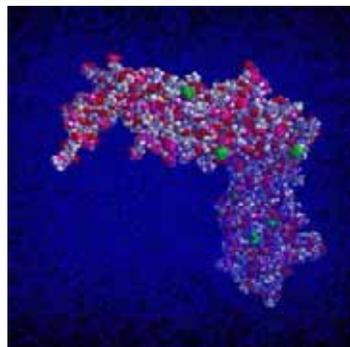


50 Å

最近の典型的な系

べん毛蛋白質
Flagellin1分子

分子量41,000
全系12万原子

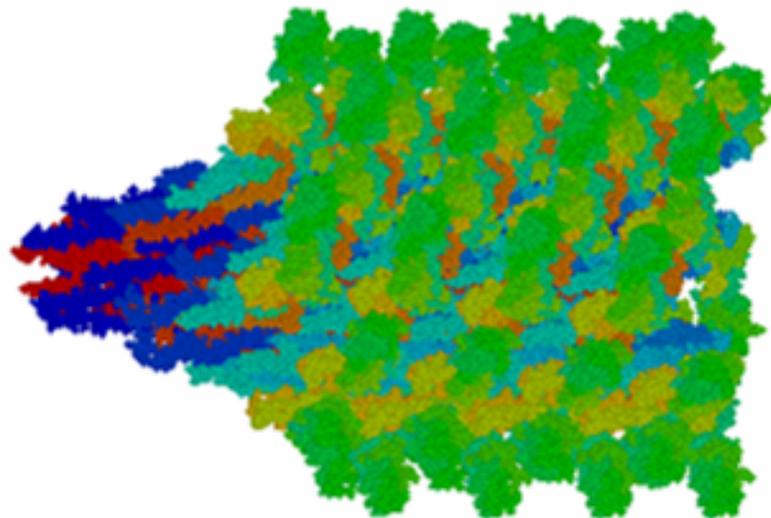


100 Å

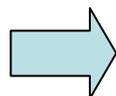
世界最大級の系

べん毛蛋白質
Flagellin44量体

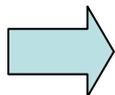
分子量240万
全系240万原子



350 Å



スパコンによる
超並列計算



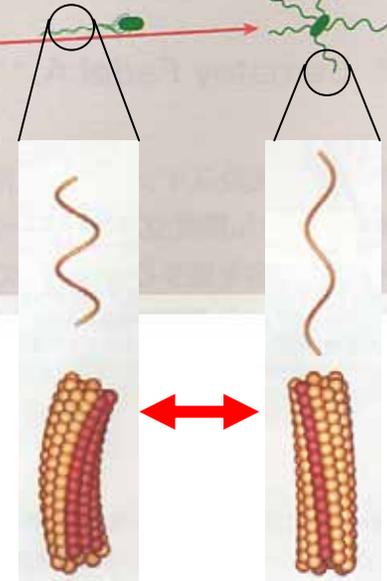
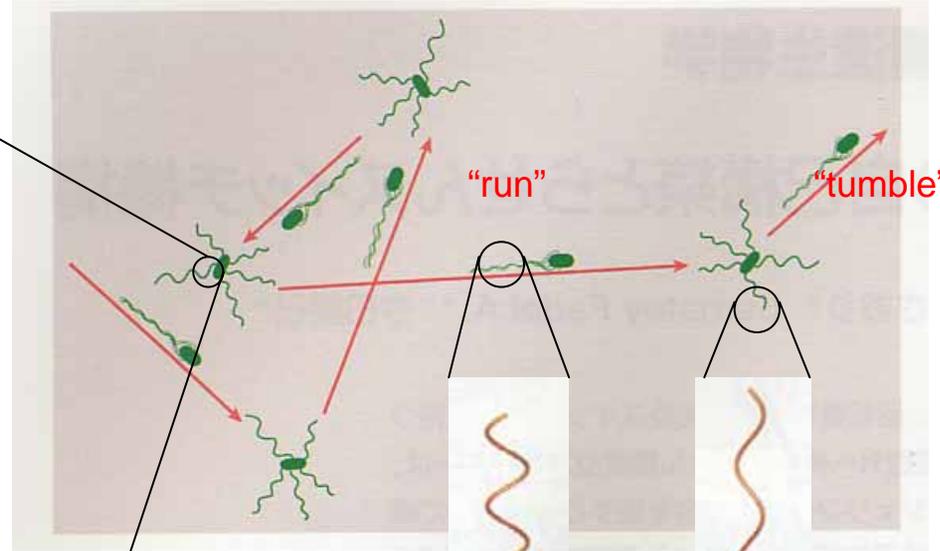
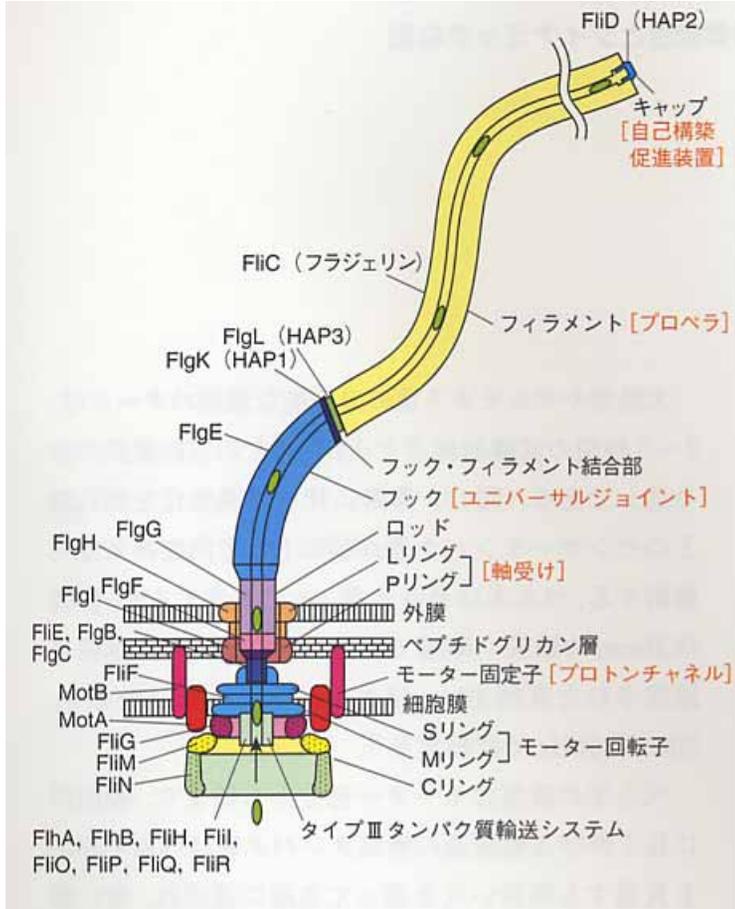
原研 ITBL計算機



べん毛による細菌の運動

べん毛複合体とべん毛繊維の超らせん構造変化

*Salmonella typhimurium*のべん毛



Normal (左巻き) Semicoil (右巻き)

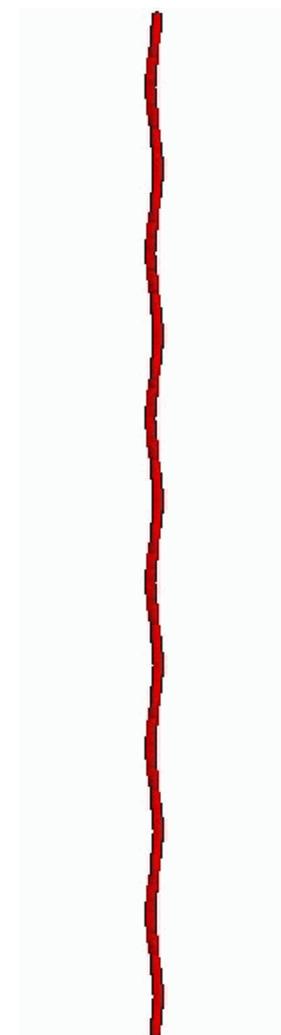
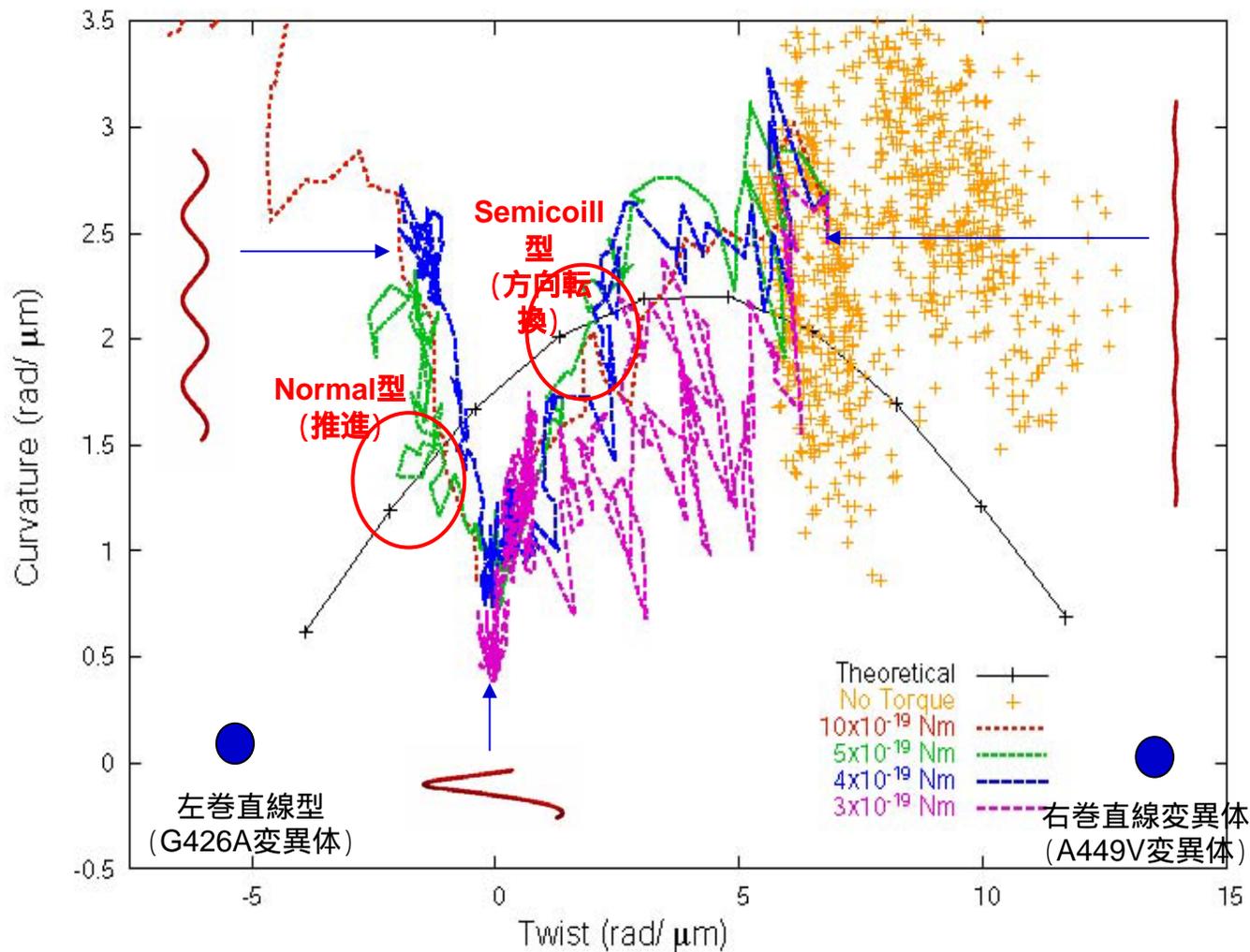
べん毛繊維(flagellin)
トルク、アミノ酸変異、塩濃度、
pHにより超らせん構造を変化

相互作用変化
による生物の
運動制御

MDべん毛の超らせん構造変化

超らせん構造変化
Torque 4×10^{-19} Nm

トルクによる超らせんパラメタの変化

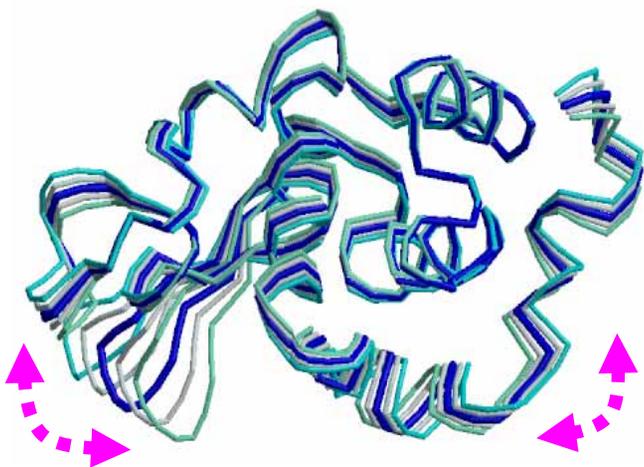


東大分生研:
北尾彰朗

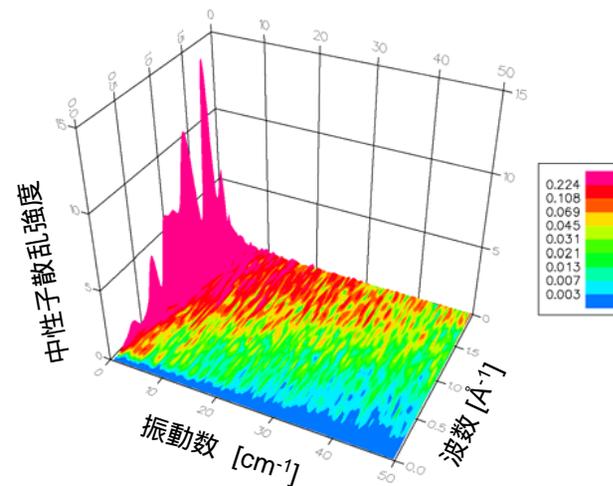
分子シミュレーションによる 蛋白質の中性子散乱スペクトル解析

蛋白質の分子シミュレーション結果から
中性子非弾性散乱データ $S(Q, \omega)$ を直接計算

分子シミュレーション



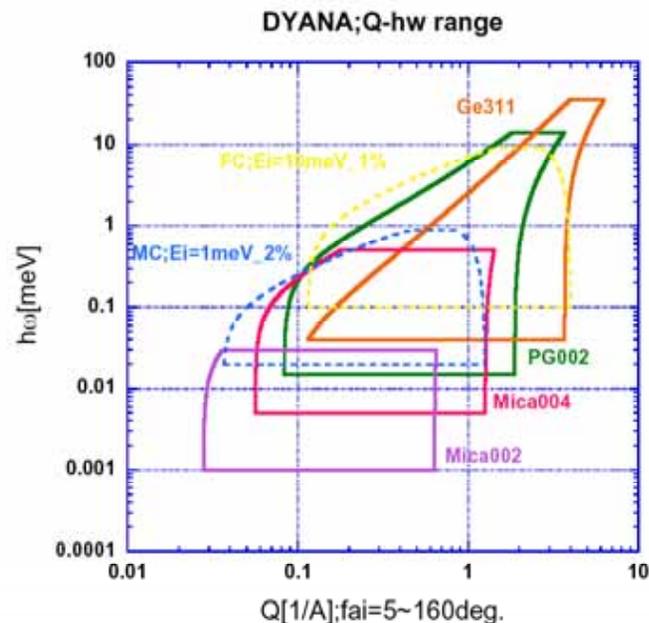
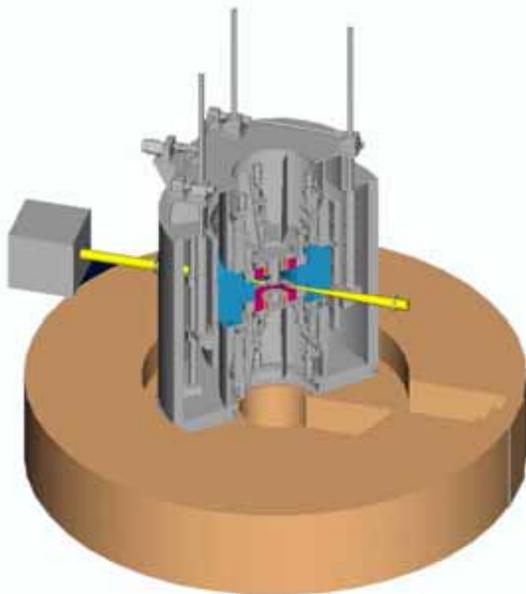
中性子非弾性散乱



分子シミュレーションを利用して
生体高分子の中性子散乱実験の可能性を探る。

(東大分生研: 城地保昌、原研中性子: 徳久淳師、奈良先端大: 中川洋)

生物用非弾性散乱装置「DYANA」



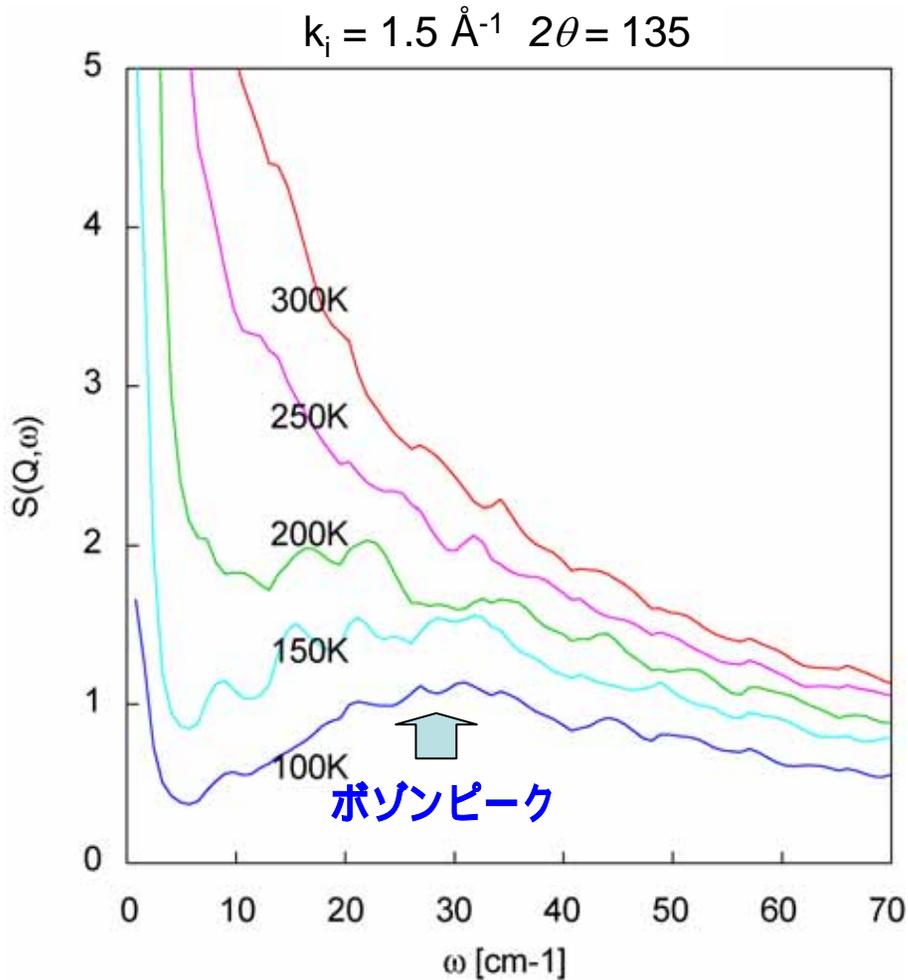
[特徴]

- 高エネルギー分解能 ($\Delta E/E_f \sim 0.5\%$)
- 高検出効率 (同一線源でIRISの約80倍以上、1MW運転時約500倍以上)
- 広いエネルギー測定範囲 (~4.5 桁)
- 偏極解析の容易さ(干渉性、非干渉性散乱の分離)

解析結果は、生物用非弾性散乱装置「DYANA」を、J-PARC物質生命科学
研究施設に建設提案申請する際に生かされた。

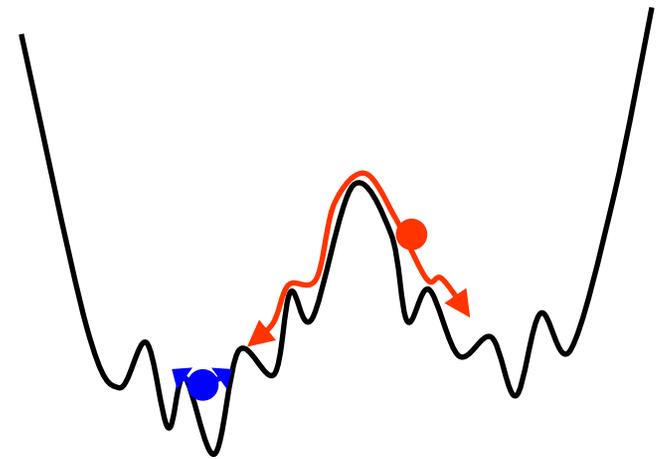
(原研中性子装置開発: 柴田薫、東大分生研: 城地保昌)

研究例：蛋白質のボゾンピーク解析



MD in water

蛋白質のボゾンピークは
構造水によって生み出される
エネルギー面の微細構造に由来する



(東大分生研:城地保昌、北尾彰朗)

低分解能生体超分子像からの原子構造構築技法

研究代表：由良 敬
研究メンバー
川端 猛 (奈良先端)、石田 恒 (原研)、
松本 淳 (原研)、真柳浩太 (BERI)、岩崎憲治 (阪大)

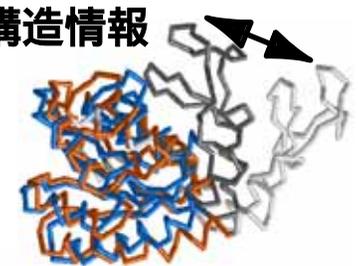
電子顕微鏡で得られる低分解能生体超分子像から
高分解能分子構造を構築する。

- ・ バイオインフォマティクスを導入し、要素タンパク質の構造推定と、要素タンパク質間相互作用部位推定を行う。
- ・ 生体分子シミュレーションを導入しタンパク質の構造変化に対応する。

分子進化情報



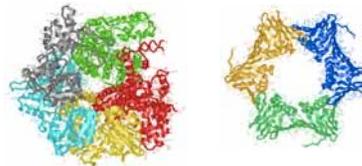
動的構造情報



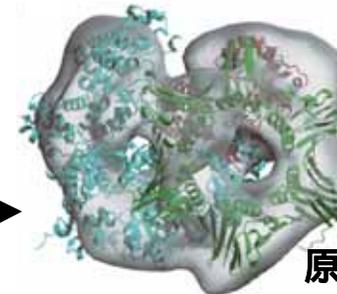
電顕顕微鏡像



X線結晶解析像



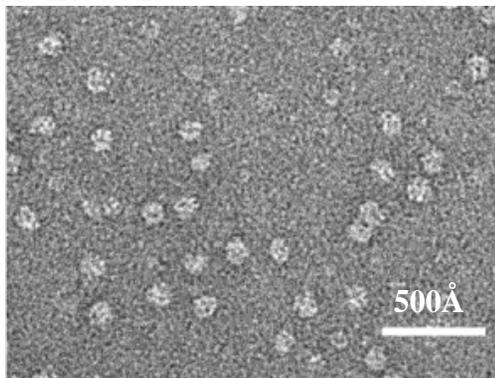
原子座標



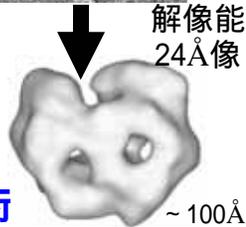
生体超分子構造は電子顕微鏡とX線結晶解析によって明らかにされている

例) 8個の要素タンパク質から構成される生体超分子

電子顕微鏡による生体超分子構造決定

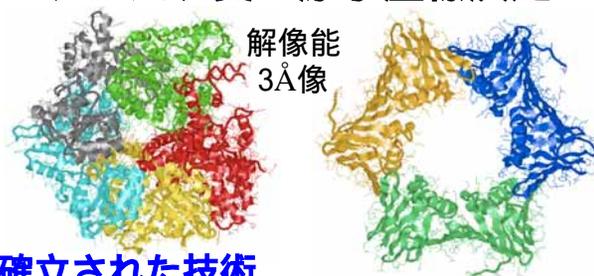


画像処理による
3次元再構成



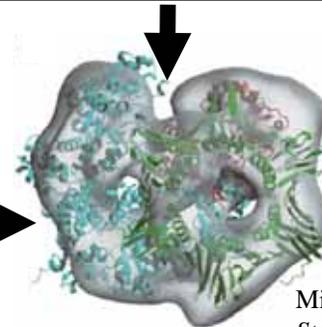
ほぼ確立された技術

X線結晶解析による要素タンパク質の原子座標決定



確立された技術

原子座標の電子顕微鏡像へのあてはめ
必要な新技術



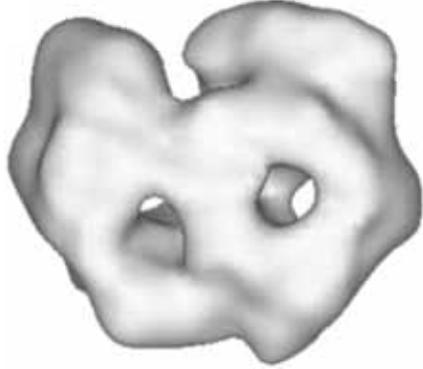
Miyata, T. *et al.* (2004) *Nature Struct. Biol.* **11**, 632.

画像認識の手法を用いた既存あてはめ法の問題点

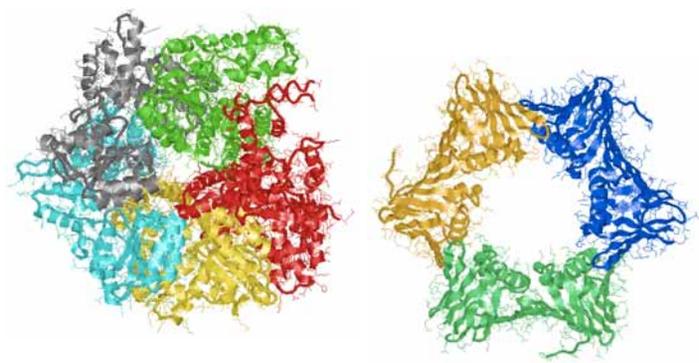
1. 要素タンパク質の構造変化に対応できない
2. あてはめ結果が一意的に決まらない

シミュレーションとバイオインフォマティクスを適用した超分子構造構築をめざす

電子顕微鏡によって得られる
低解像度生体超分子構造

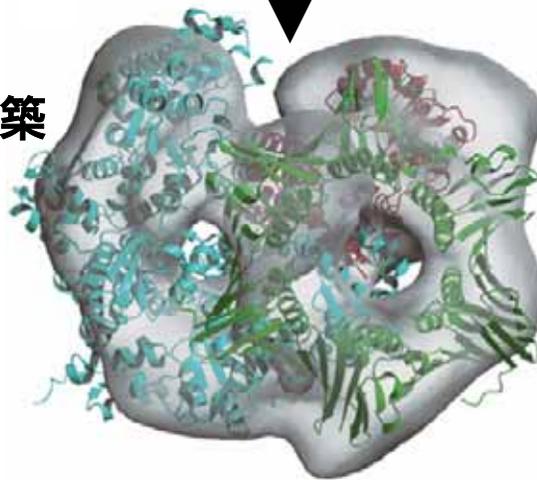


X線結晶解析によって得られる
高分解能要素タンパク質構造



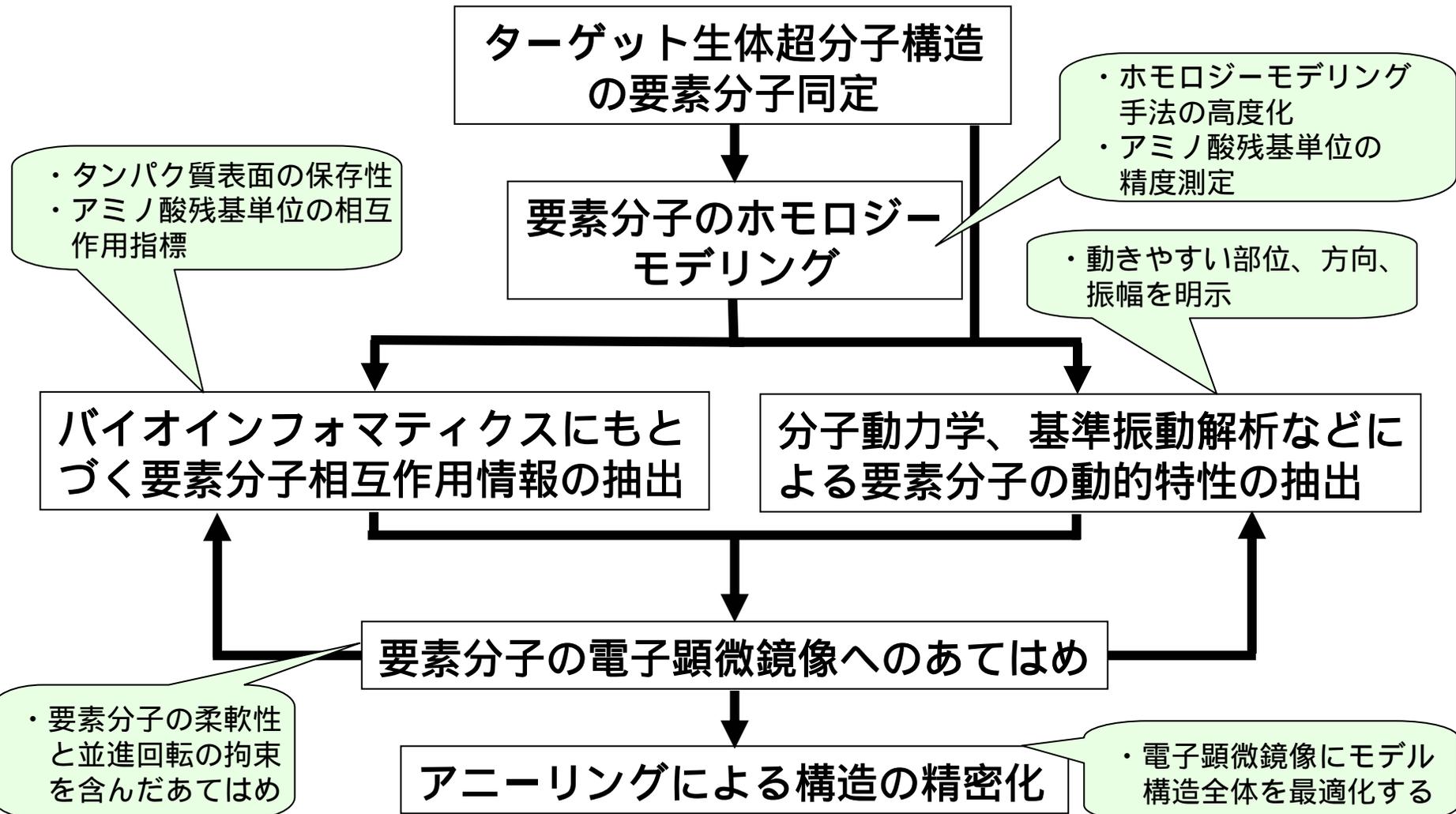
生体分子シミュレーションと
バイオインフォマティクスを導入

生体超分子の
原子解像能構造構築



奈良先端大学院大学
大阪大学超高压電子顕微鏡センター
生物分子工学研究所構造解析研究部
との協力研究 (CREST)

生体分子シミュレーションとバイオインフォマティクスを導入



要素分子 = 要素タンパク質/DNA/RNA