

バイオシミュレーション研究者の会 平成17年度研究報告

平成18年1月6日

バイオシミュレーション研究者の会代表

高田 俊和

(代理: 岡本 祐幸)

“バイオシミュレーション研究者の会”活動目的

次世代の重要分野として、バイオテクノロジーが注目される中、

- 1) 生体機能の発現メカニズムを、分子レベルで明らかにする。
- 2) それらの知見・技術を、創薬など産業上の活用に結びつける。

ための分子シミュレーション技術構築とプログラム開発を行い、有効性を示すための実証計算を行う。

“バイオシミュレーション研究者の会”研究課題

サブテーマ1:「蛋白質の高次構造変化のリアルなシミュレーション」

斎藤稔 弘前大学・理工学部

サブテーマ2:「第一原理からのタンパク質の折り畳みシミュレーション」

岡本祐幸 名古屋大学・理学部

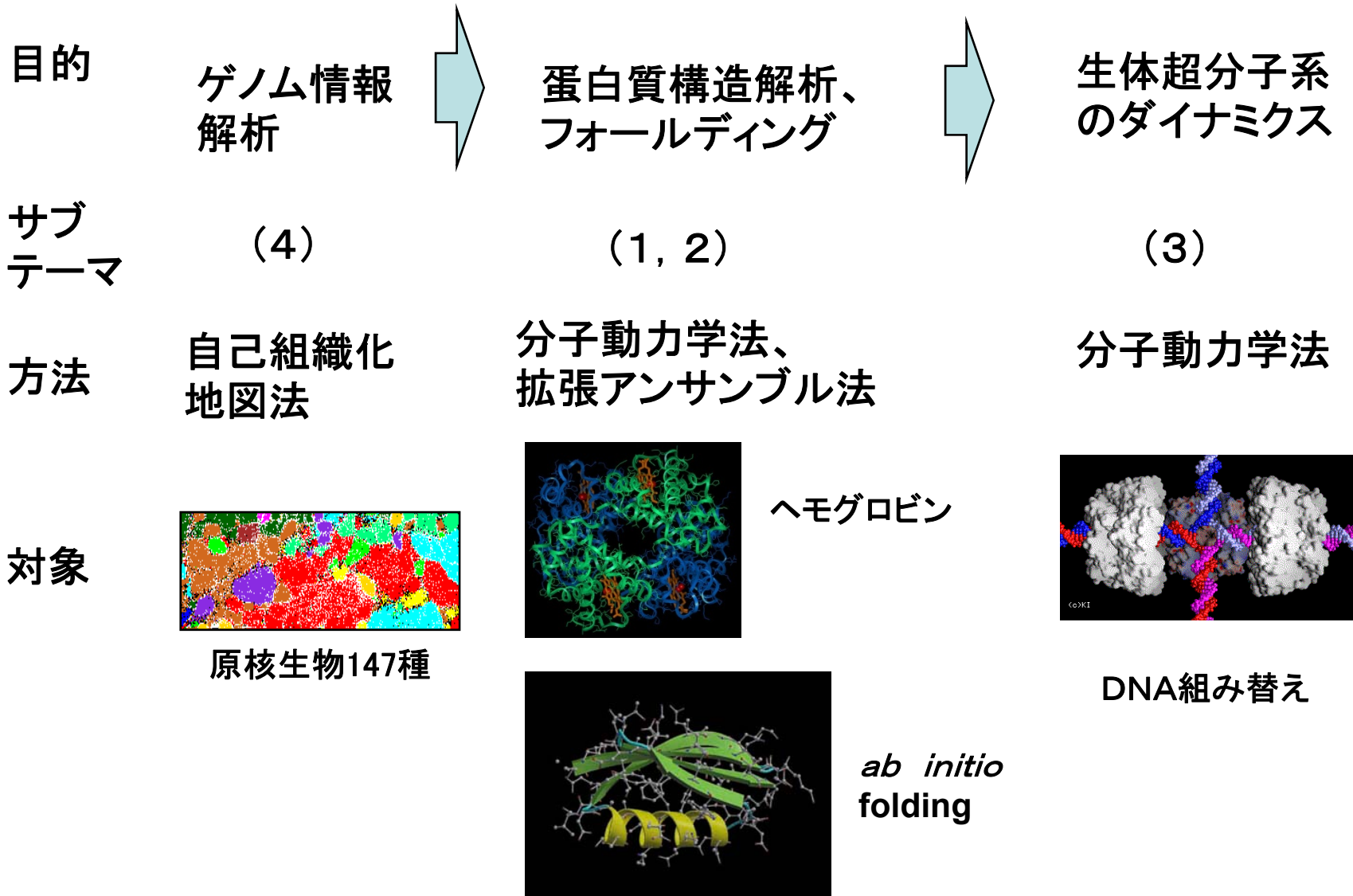
サブテーマ3:「分子動力学シミュレーションを用いた大規模生体超分子系の機能解析」

石田恒 日本原子力研究所・計算科学技術推進センター

サブテーマ4:「全ゲノム配列と全タンパク質配列の自己組織化地図作成と進化シミュレーション」

池村淑道 総合研究大学院大学・葉山高等研究センター

サブテーマの研究内容と相互関係



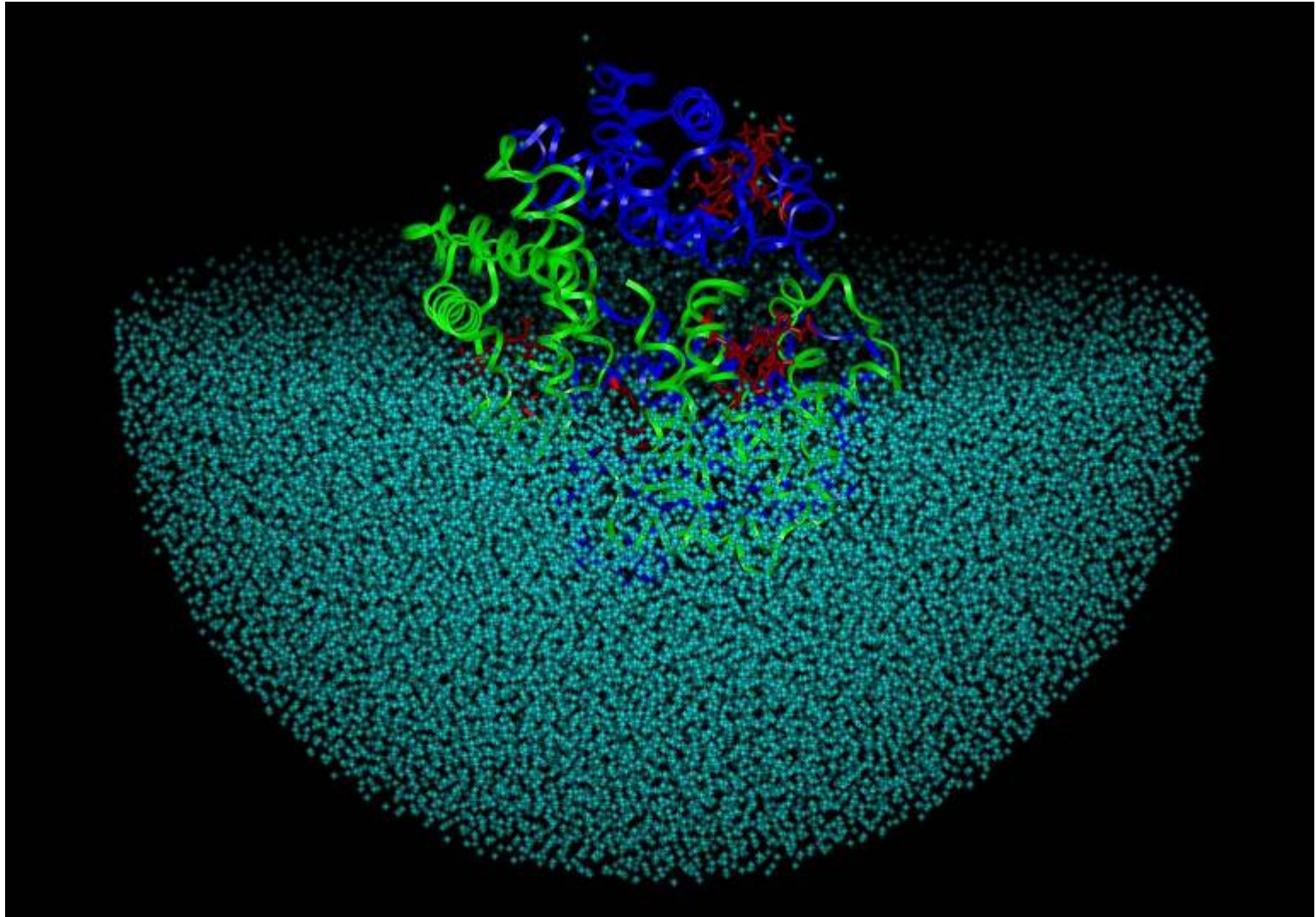
蛋白質の高次構造変化の リアルなシミュレーション

斎藤稔、岡崎功(弘前大学)

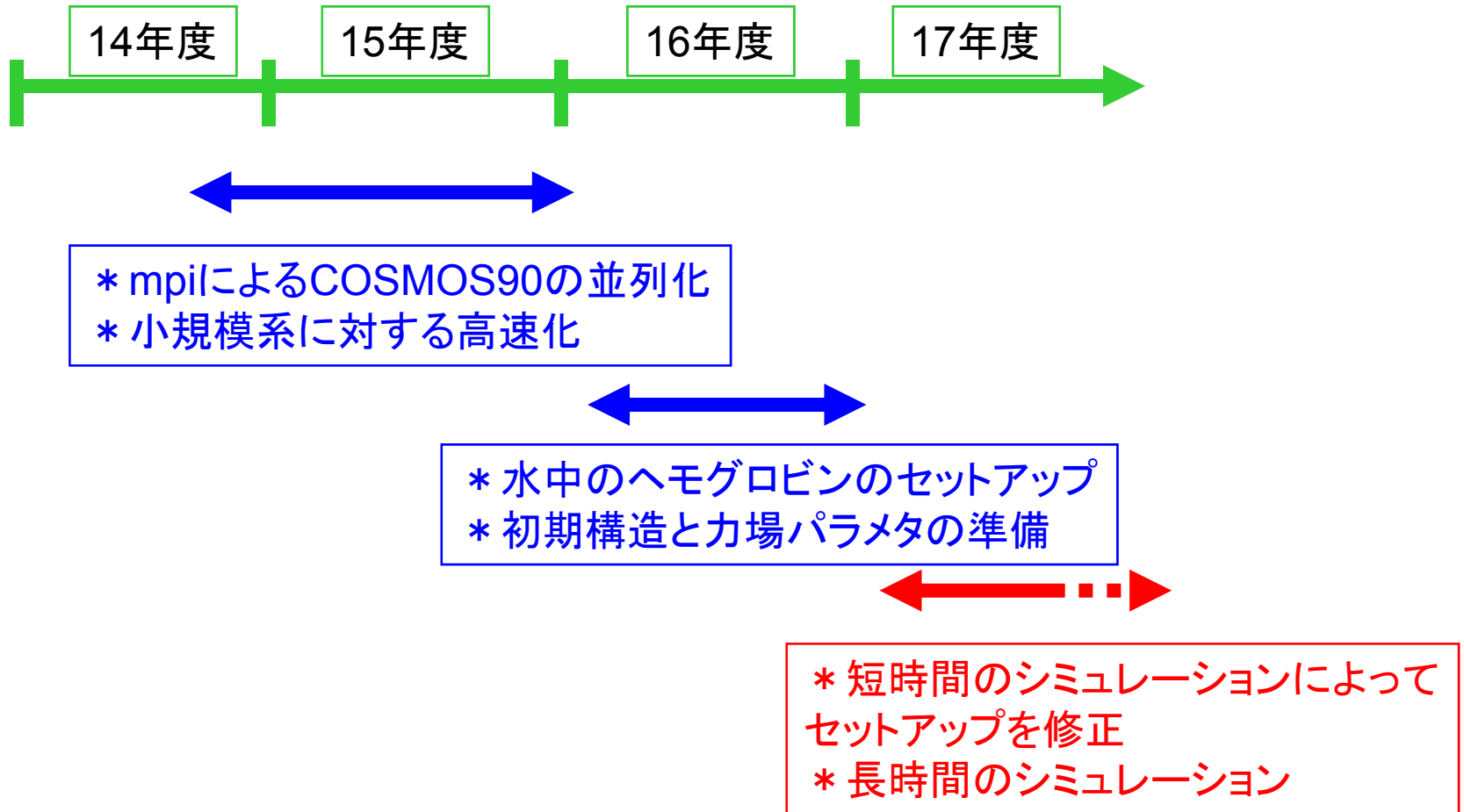
ヘモグロビンの酸素吸着に伴う4次構造変化の過程は、実験によってとらえることができない。そこで、コンピュータシミュレーションによって構造変化を追跡し可視化する。

水中のヘモグロビン

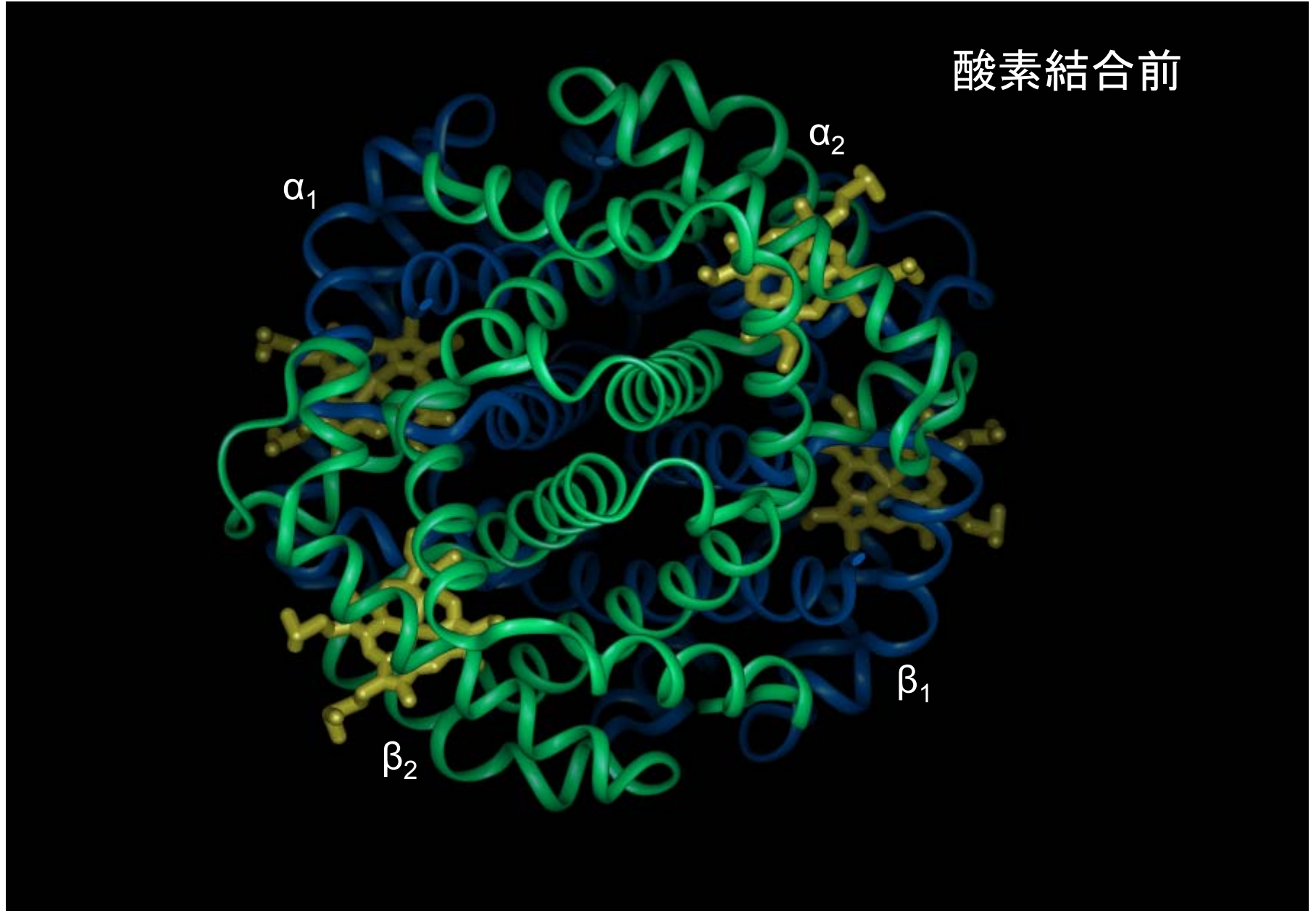
大規模でリアルなシミュレーション



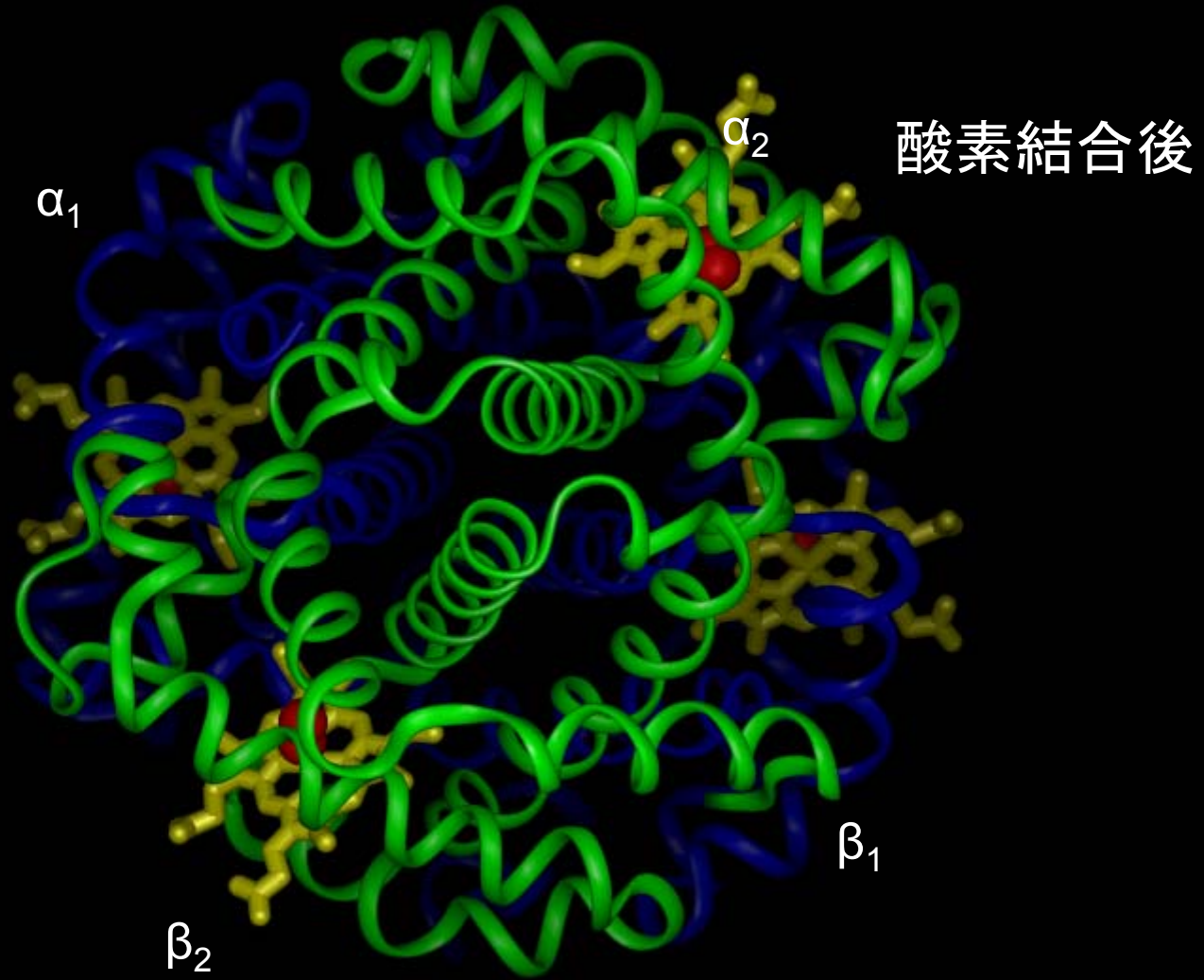
研究の進捗状況



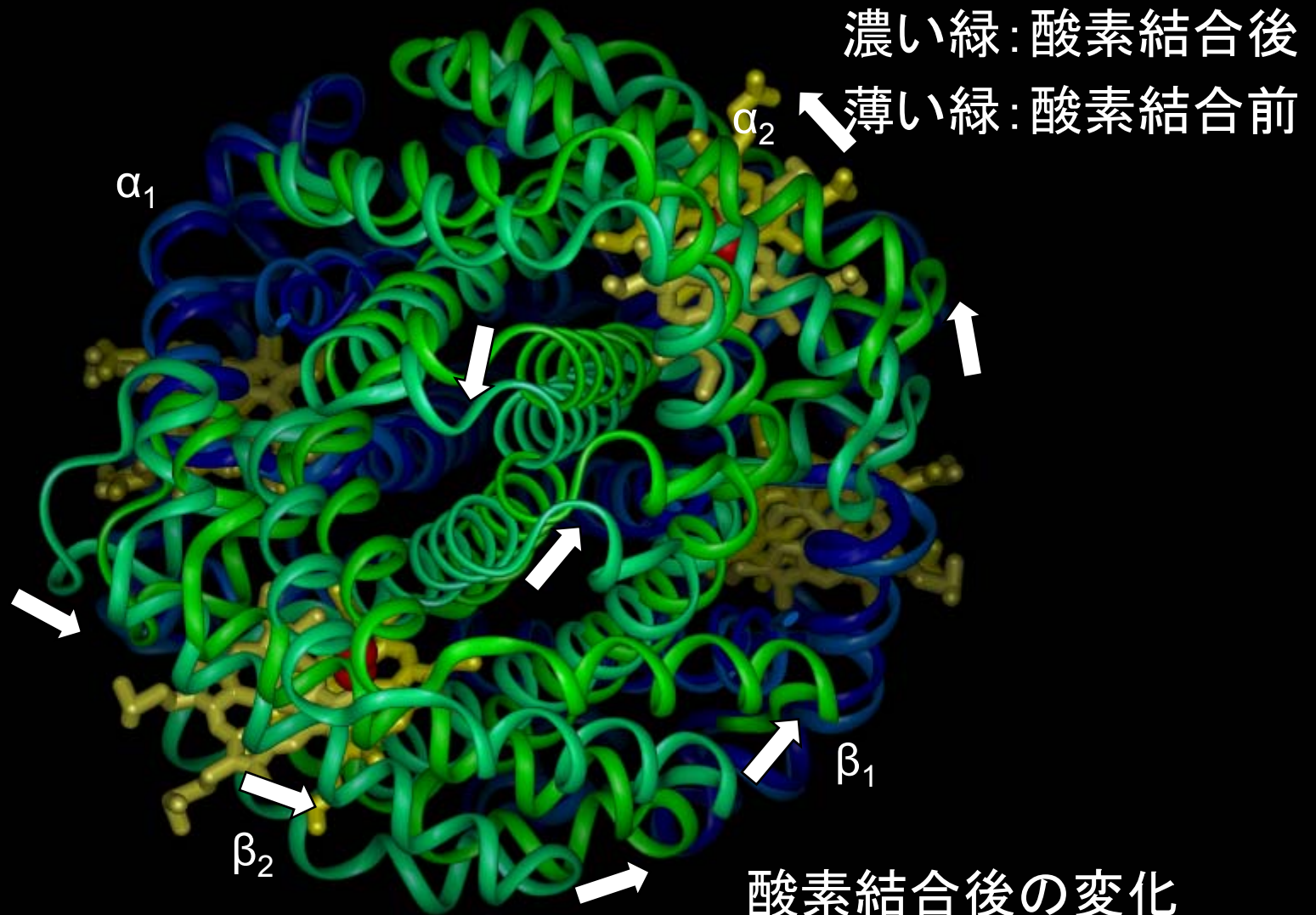
ヘモグロビンの4次構造変化(実験事実)



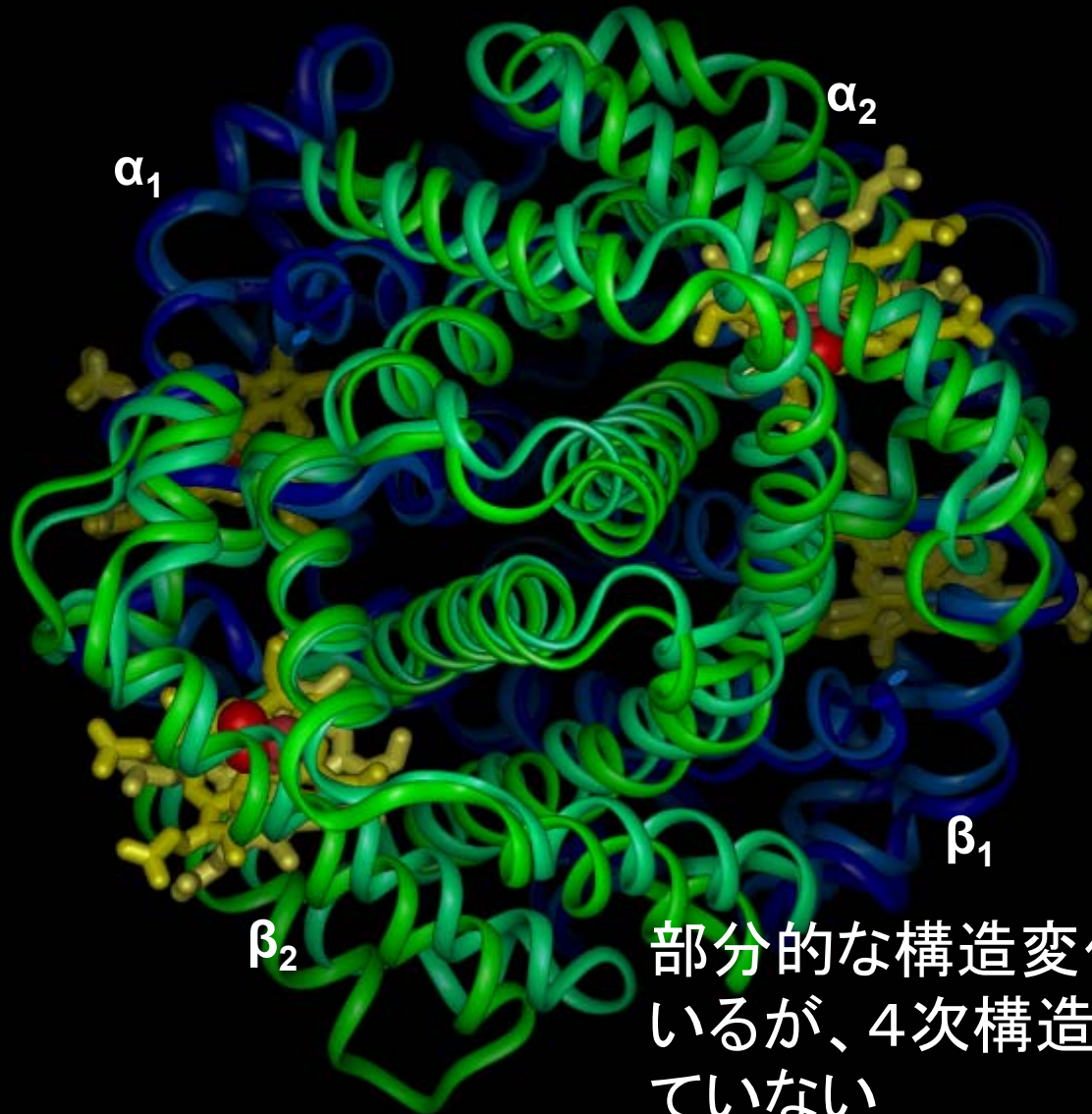
ヘモグロビンの4次構造変化(実験事実)



ヘモグロビンの4次構造変化(実験事実)

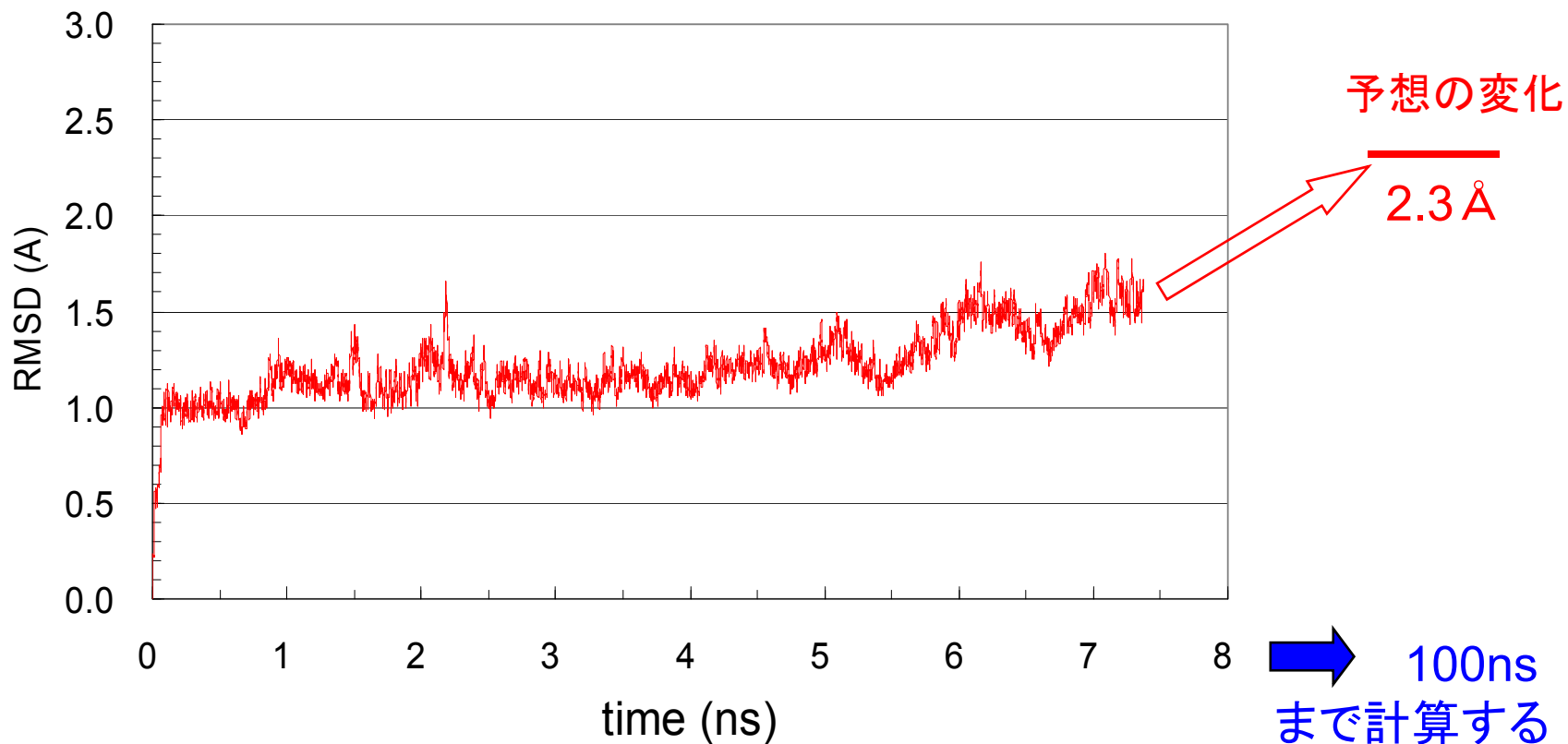


現在までの構造変化(シミュレーション)



初期構造からの変化

RMSD(主鎖の最小二乗変位)



シミュレーションは安定している
ネットワークからジョブを投入して長時間のシミュレーションを引き
続き行う

第一原理からのタンパク質の 折り畳みシミュレーション

岡本祐幸(名古屋大学)

光武亜代理(慶応大学)

杉田有治(東京大学)

依田隆夫(長浜バイオ大学)

西川武志(産総研、東京工業大学)

榮慶丈(広島大学)

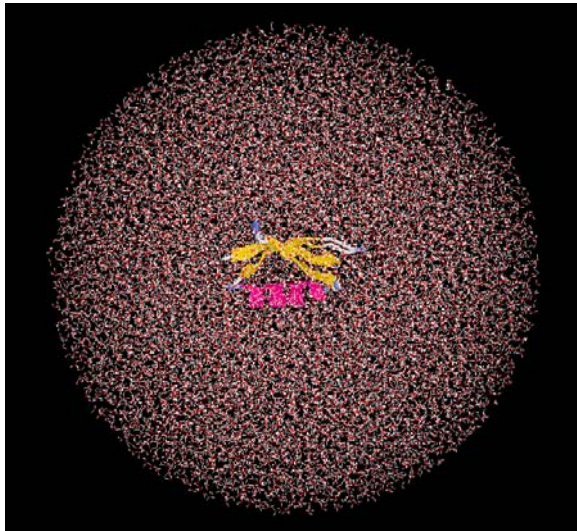
水中のタンパク質の折り畳みシミュレーション

研究計画

半径50 Åの水球中(水分子17,063個)の小タンパク質 protein G(アミノ酸数56、全原子数52,044個)を拡張アンサンブルシミュレーションにより折り畳む。

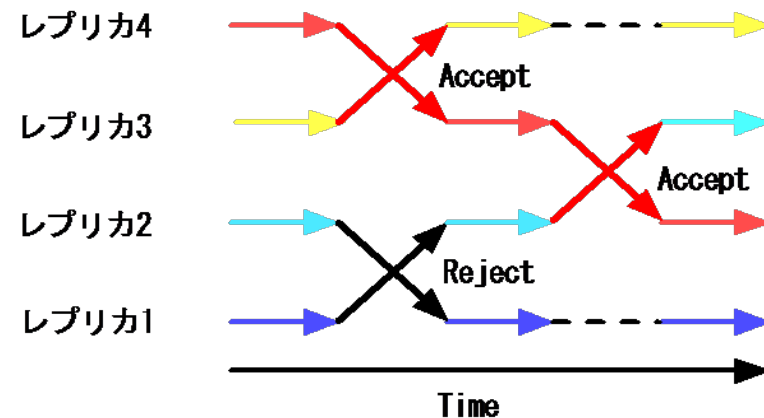
拡張アンサンブル法(Generalized-Ensemble Algorithms) エネルギー空間上のランダムウォークを実現することにより、エネルギー極小状態に留まるのを避ける。

[拡張アンサンブル法の例:レプリカ交換法]



異なる温度を持つ系のコピー(レプリカ)をM個用意(M=224)

1. 各レプリカで独立にカノニカルシミュレーションを実行
2. 途中で2つのレプリカの温度を交換
この2つの操作を繰り返す



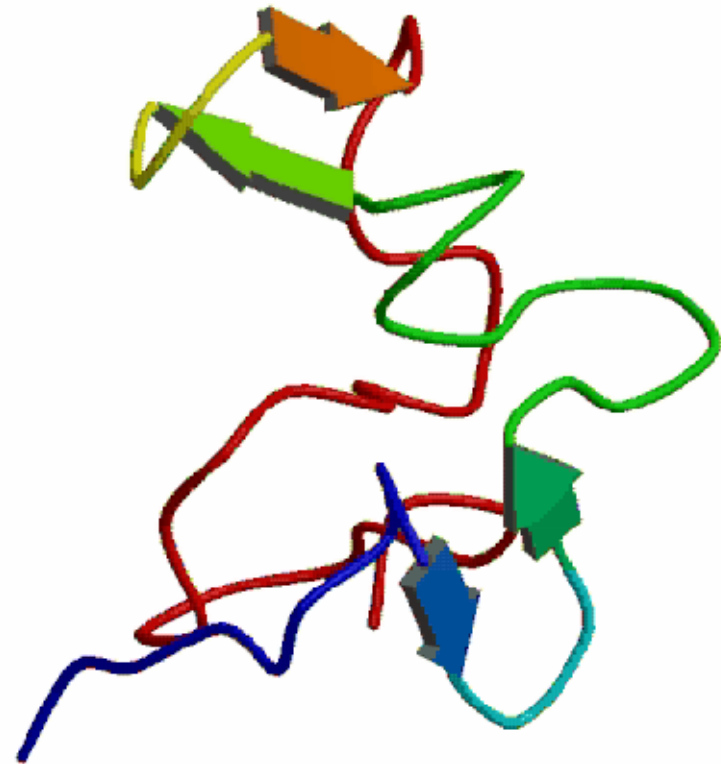
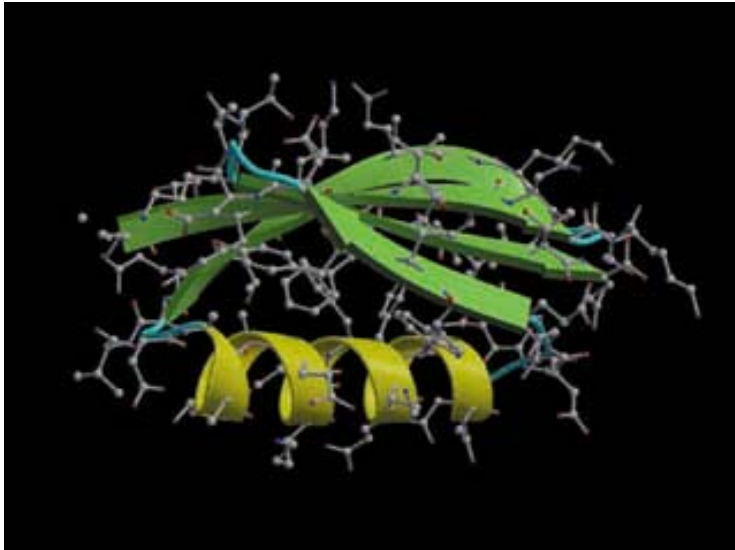
水中のタンパク質の折り畳みシミュレーション

レプリカ交換分子動力学シミュレーション:
(タンパク質の主鎖のみ描画し、側鎖と水分子を省略)

初期構造:完全に伸びた構造



目標:X線回折実験の構造(緑はβ
シート、黄はαヘリックス)

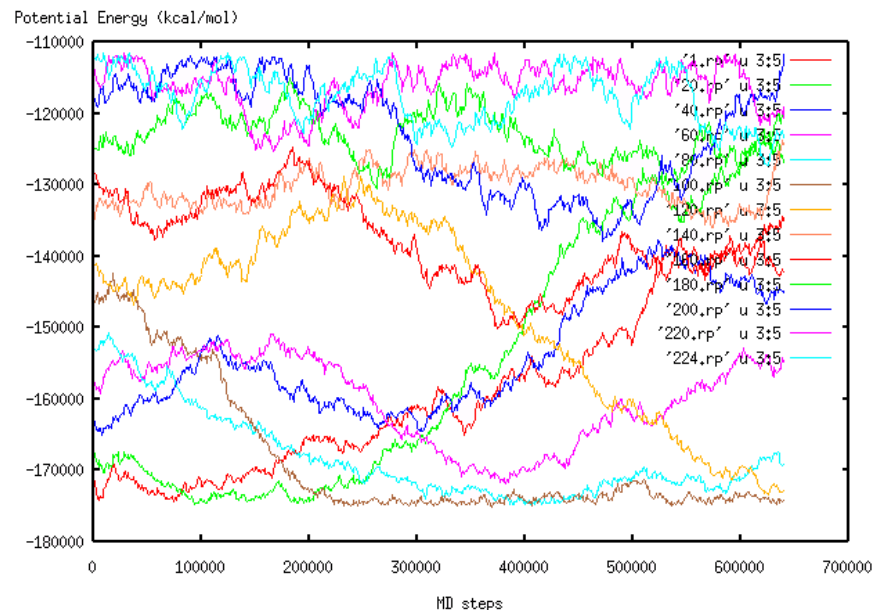


水中のタンパク質の折り畳みシミュレーション解析結果

エネルギー空間上のランダムウォークを実現
(右図)

→ 従来の手法では得られない程幅広い構造空間のサンプリングに成功

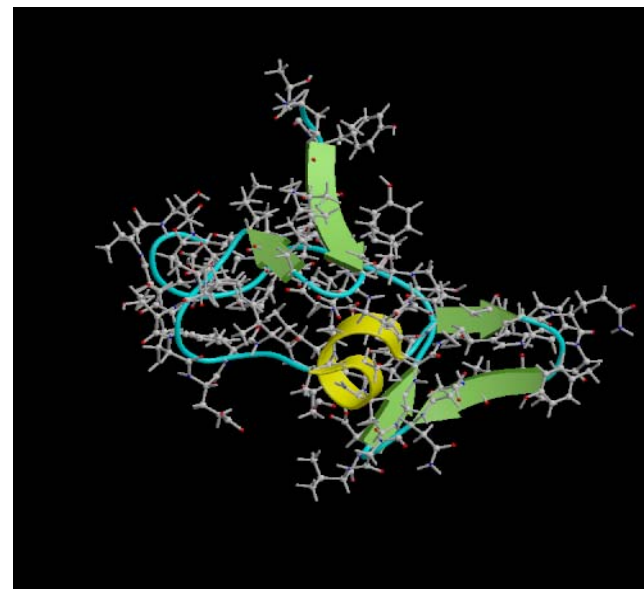
→ α ヘリックスや β シートなどの2次構造が頻繁に生成したり壊れたりするのを観測



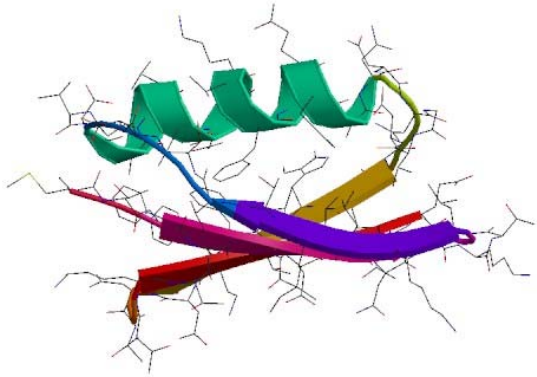
このような大規模で複雑な系における拡張アンサンブルシミュレーションに初めて成功した

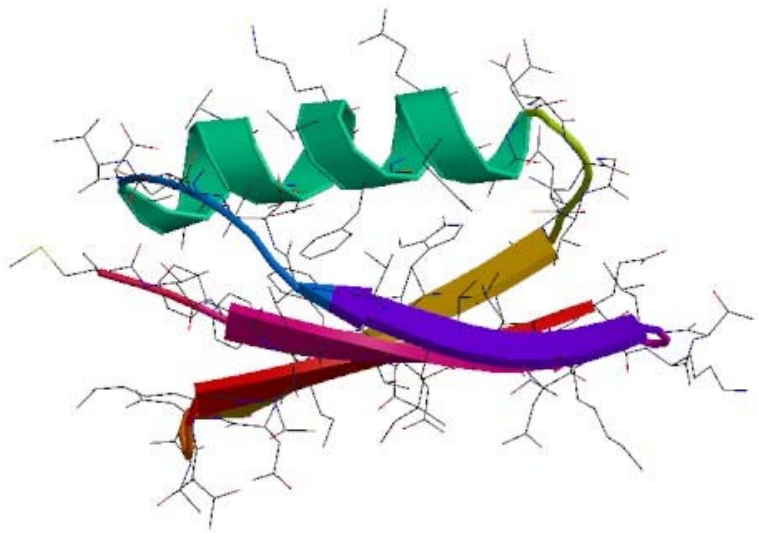
特に、完全に伸びた初期構造からシミュレーションを始め、自然の立体構造と似た場所に2次構造(α ヘリックスや β シート)の形成が観測された
(右図参照)

まだ、計算時間が十分でないので、自然の構造への折り畳みは達成されていないが、その達成のための準備を整えているところである



224レプリカのREMDシミュレーション(112ノード使用)
を各レプリカにおいて **5 nsec** (よって、トータルで
1.12 μ sec)まで延長





分子動力学シミュレーションを用いた 大規模生体超分子系の機能解析

石田恒(日本原子力研究所)

北尾彰朗(東京大学)

城地保昌(東京大学)

樋口真理子(日本原子力研究所)

米谷佳晃(日本原子力研究所)

叶野琢磨(日本原子力研究所)

松本淳(日本原子力研究所)

巨大生体高分子のシミュレーションシステム

SCUBA (Simulation Codes for hUge Biomolecular Assembly) 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

大規模化、高速化

スーパーコンピュータ

生体高分子複合体系
10⁶粒子、10⁶ステップ以上

SCUBAの高速化・高精度化

並列化

空間分割法による通信コスト削減

長距離相互作用

PPPM (Particle-Particle Particle-Mesh)法を採用

並列化率の向上

動的ロードバランスによる各CPU負荷の均一化
並列化率の異なる計算(PP計算とPM計算)をCPU上で切り分け

ベクトル化率の向上

PP計算 (長距離相互作用の実空間計算) のベクトル化強化

SCUBAの計算性能

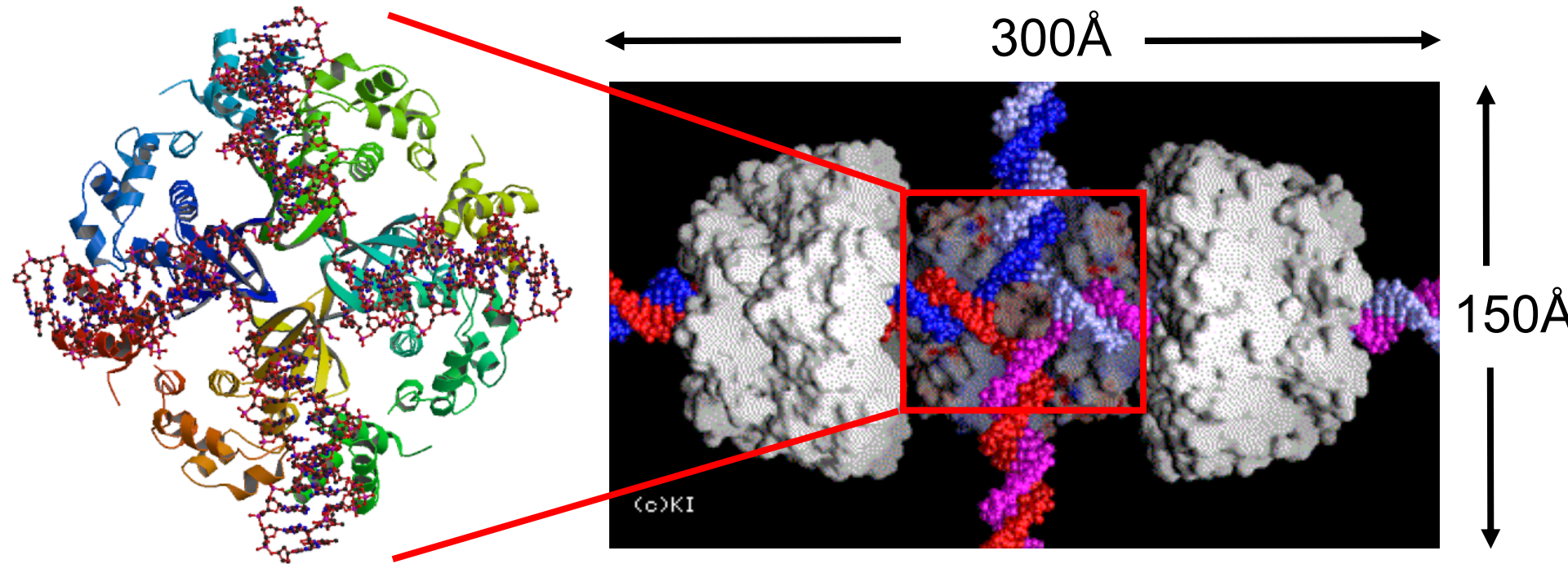
45ノード(360CPU)による最適計算
並列化効率75.9%
ベクトル化率96.2%



大規模生体超分子系 (RuvA-RuvB-Holliday分岐DNAの系、約50万原子) でMDシミュレーションを実行中

生物学的意義：DNA組換えのダイナミクス解明

DNA組換えのモデル

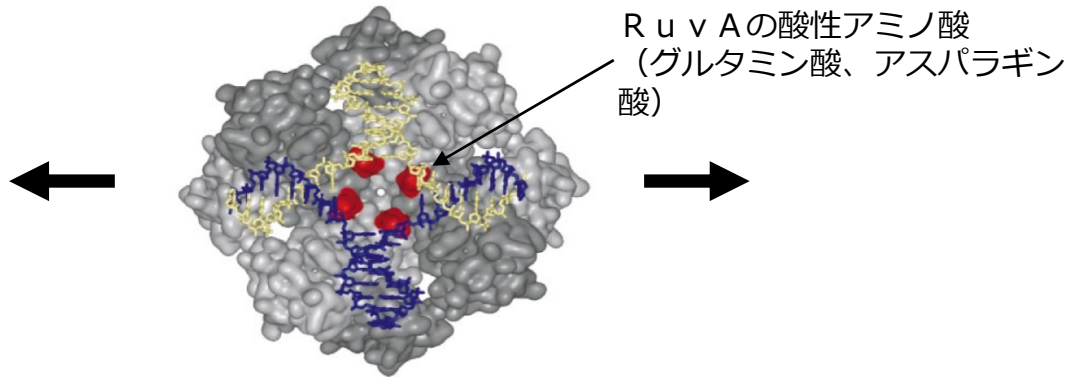


DNAに結合したRuvA蛋白質
(本年度解析)

DNAに結合したRuvAB蛋白質
(現在計算中)

DNA組換えのシミュレーション

DNAが組み換わる様子を追跡



DNAに結合したRuvA 4量体蛋白質の系

111,224原子 (RuvA: 12,628 原子, DNA: 3196 原子, 水: $31,764 \times 3$ 原子, Naイオン: 108 原子)

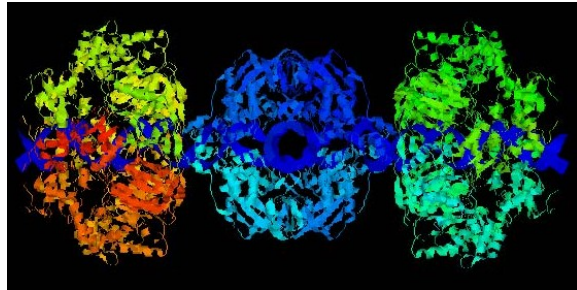
見積もられたDNA組換えの自由エネルギー
A-T、G-Cで構成された場合、約10-15kcal/mol

DNA組換えの中間体では、RuvAの酸性アミノ酸と組換え塩基との間に安定な相互作用が観測された。

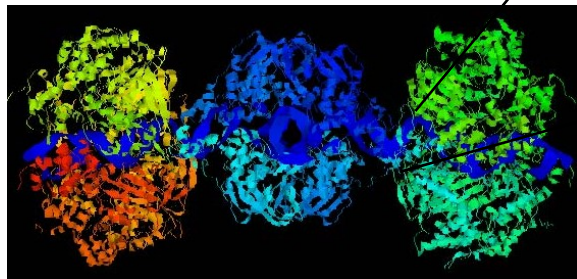
組換えDNAの遷移:

DNAの水素結合長、
組替え塩基と酸性アミノ酸との
相互作用などを詳細に解析中

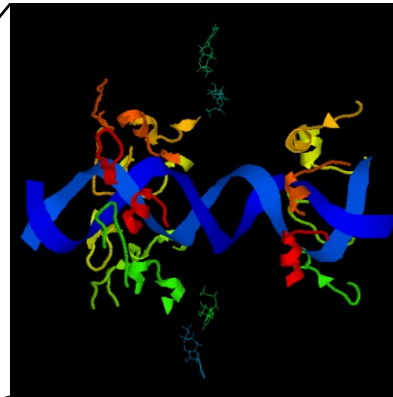
モータータンパク質を含めたDNA組換えのシミュレーション (現在計算中)



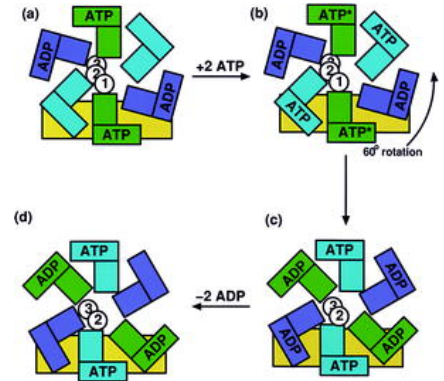
RuvA-RuvB-DNAの
初期構造モデル(約50万原子)



MDシミュレーション平衡化後の
RuvA-RuvB-DNAの構造



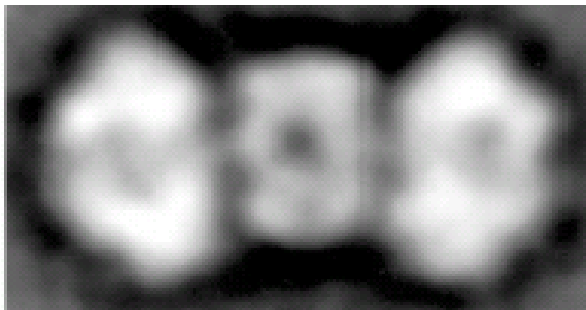
RuvBとDNAのコンタクト部位



RuvBがATP加水分解エネルギーを利用して
DNAを移動させるモデル

1. SCUBAで超大規模計算(約50万原子)が可能であることを確認した。(45ノード計算で50%以上の並列化効率、95%以上のベクトル化率を達成)
2. RuvA-RuvB-DNAの平衡状態における構造は、実験で観測された構造と似ている。
3. モーター蛋白質RuvB6量体のサブユニットにおいて、ATPの結合状態により、構造の違いが見られた。

今後:モーター蛋白質RuvBのどのようなダイナミクスがDNAを移動させるのかを調べる。



RuvA-RuvB-DNAの電子顕微鏡像

全ゲノム配列と全たんぱく質配列
の自己組織化地図作成と
進化シミュレーション

池村淑道(総研大)

阿部貴志(遺伝研)

大量かつ複雑な情報から似かよった情報が自己組織化する (おのずと集まる)

自己組織化地図法 (Self-Organizing Map : SOM)

配列の特徴だけを基にゲノム断片の
高精度な分類が可能

難培養性微生物のゲノム解析への有用性

Abe, T. et al. (2003)

Informatics for unveiling hidden genome signatures.

Genome Res., 13(4), 693-702 .

ゲノム配列に潜む生物種の個性のSOM解析

生命の設計図であるゲノムは、4種類の文字(A, T, G, C; 塩基と呼ぶ)で書かれている。

ACAGATTAGACCCTGAC-----

例えば、ヒトゲノムの場合は、30億文字(3Gb)で書かれており、朝刊の新聞に例えると、25年分。

塩基配列が既知なすべての生物のゲノム配列を対象に、各々を1万文字(10 kb)に断片化して以下の単語を数える。

- | | |
|----------------------------------|------------------|
| 2連塩基: AA, AC, ----- : | 16種類 の単語 |
| 3連塩基: AAA, AAC, ----- : | 64種類 の単語 |
| 4連塩基: AAAA, AAAC, ----- : | 256種類 の単語 |

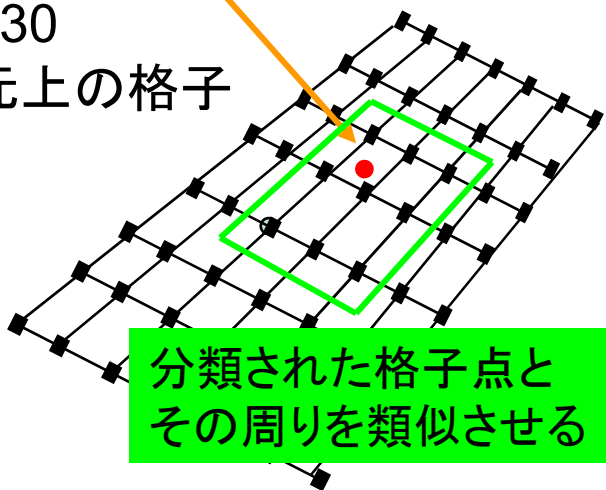
通常SOM

大統領の原稿10,000件

●, ▲, □, △, ■,

一番近いところを探す

30 x 30
2次元上の格子



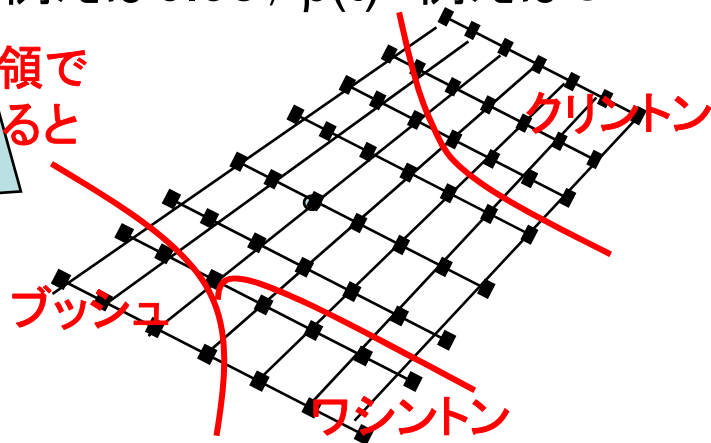
分類された格子点と
その周りを類似させる

例えば20人の大統領の各500件で、合計10,000件の演説原稿について、1000種類の単語 (Science, War, Iraq, Peace, ..) に着目したSOMを考える。

初期状態としては、大統領とは無関係の900件の原稿を集めて、各原稿を任意の順番で、30 x 30の2次元の格子に並べた上で、1000種類の単語の使用頻度を計算する。

$\alpha(t)$ 例えば0.95; $\beta(t)$ 例えば5

学習後、大統領で
結果をみると



- 入力データの順番に依存しない
- 解析の再現性

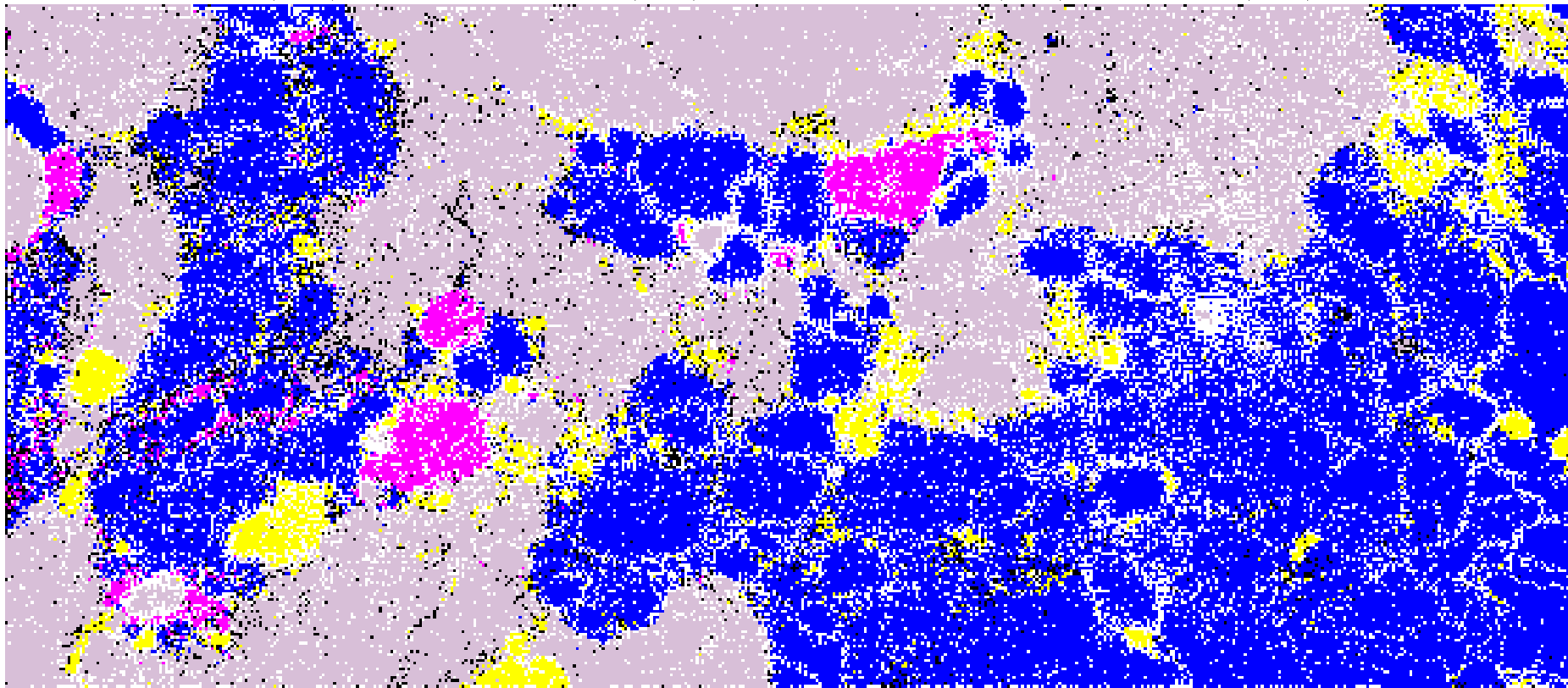


ゲノム配列解析に適応

・初期状態にランダム値から主成分分析の第1軸と第2軸へと変更

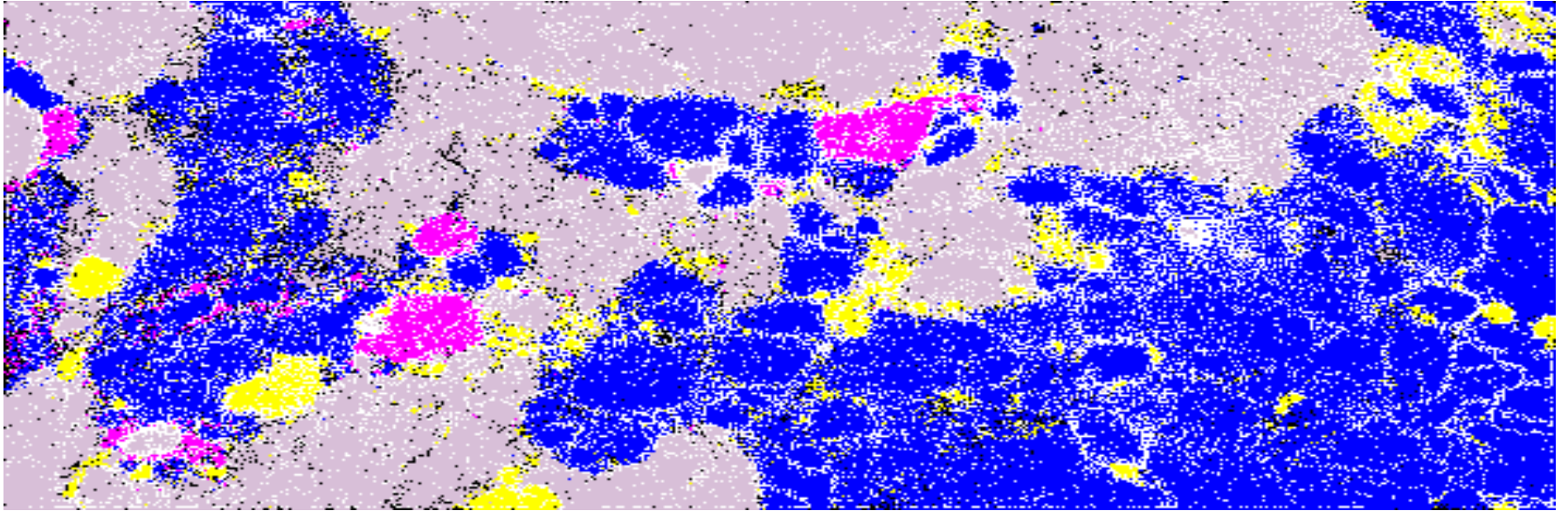
現在、入手可能な全ゲノム配列を対象としたSOM解析

真核生物(■), 原核生物 (■), オルガネラ(■), Virus (■)



真核生物40種, 原核生物1,502種, オルガネラ684種(ミトコンドリア 642種、葉緑体 42種), ウィルス1,065種での断片化サイズ5 kb, 縮退4連続塩基でのSOM縮退では、ゲノム配列の相補性の影響を除去するために、相補的な配列 (例えば、AAAAとTTTT) を同一のものとし、頻度解析を行う。

真核生物 (■), 原核生物 (■), オルガネラ(■), Virus (■)



真核生物40種, 原核生物1,502種, オルガネラ684種(ミトコンドリア 642種、葉緑体 42種), ウィルス1,065種での断片化サイズ5 kb, 縮退4連続塩基でのSOM

各カテゴリー領域へ分類されていた割合

	Eukaryotes	Mitochondria	Chloroplast	Virus	Prokaryotes	全配列数
Eukaryotes	95.58%	0.19%	0.20%	1.59%	2.44%	180,866
Mitochondria	4.88%	87.04%	3.22%	0.77%	4.10%	8,593
Chloroplast	8.57%	5.24%	72.17%	0.82%	13.19%	6,066
Virus	9.70%	0.21%	0.15%	78.50%	11.44%	32,072
Prokaryotes	1.70%	0.09%	0.35%	1.55%	96.31%	211,567

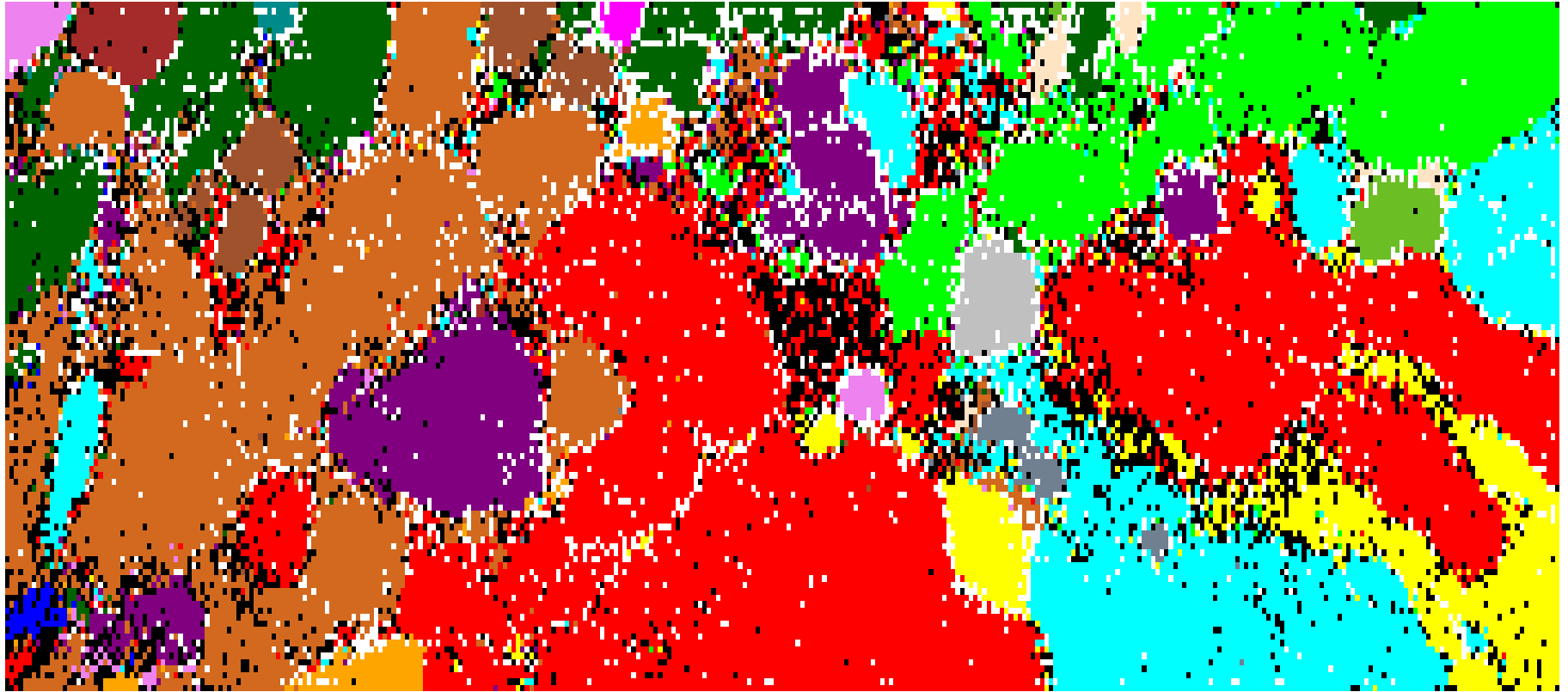
オルガネラとウィルスの分離能は、80%前後と若干分離能は低い、真核生物と原核生物については、96%と高精度に分離されている

膨大な未開拓ゲノム資源の活用

Metagenome解析

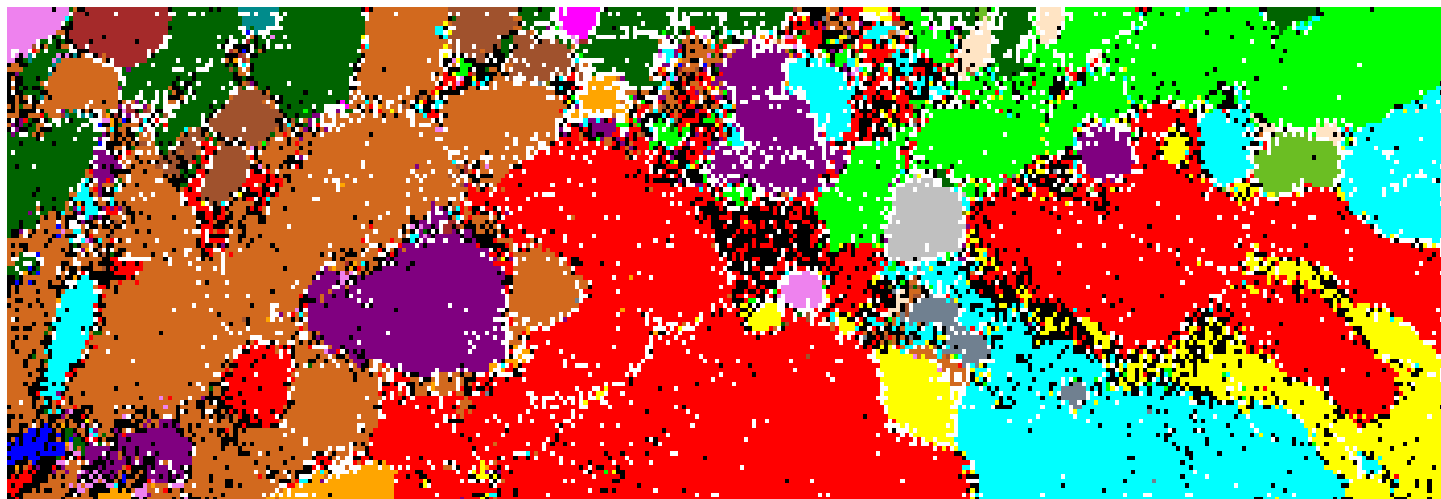
難培養性微生物のDNAを、培養することなく体内や自然環境から回収し配列決定を行い、医学や産業に重要な遺伝子を探索する

例えば、Venterは、**Sargasso Sea (near Bermuda)** の数百Literの海水由来のDNAの shotgun sequencingを行い、80万本の断片配列(約1 Gb)を得た。120万個の遺伝子の候補を推定している。**大量な短い配列(約1 kb)。**

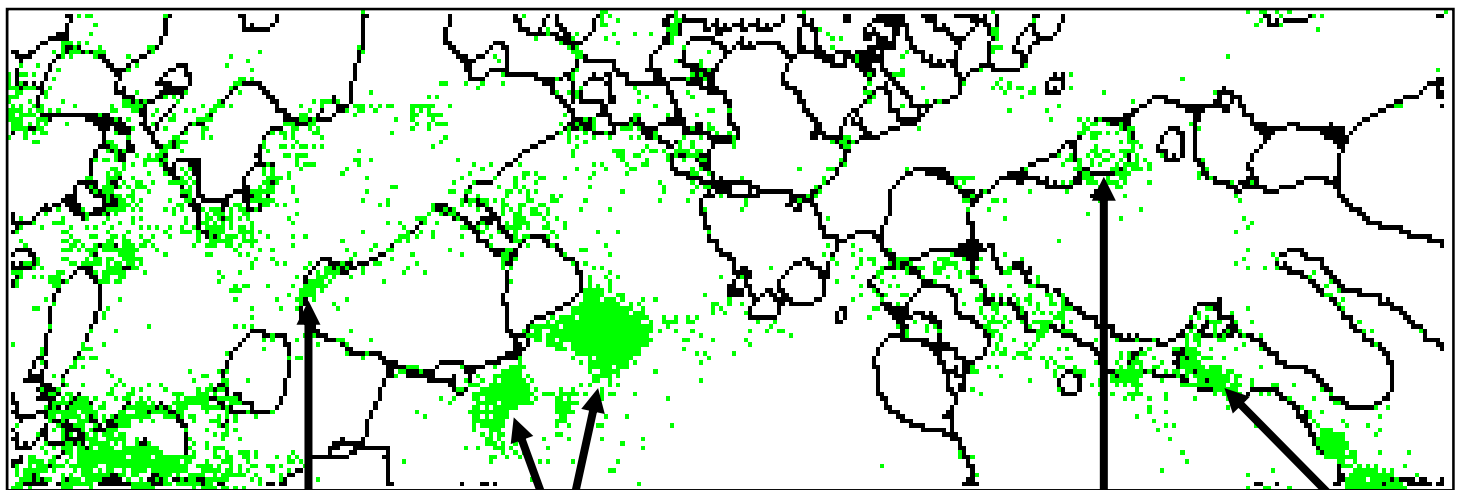


GenBankに収録されている、各ゲノムについて10 kb以上の配列が解読された全原核生物1,502種(総計約1 Gb)についての断片化サイズ5 kb配列の縮退4連続塩基のSOM(系統群にて色分け)。縮退では、ゲノム配列の相補性の影響を除去するために、相補的な配列(例えば、AAAAとTTTT)を同一のものとし、頻度解析を行う。

Actinobacteria (■), Alphaproteobacteria (■), Aquificae (■), Bacteroidetes (■), Betaproteobacteria (■), Chlamydiae (■), Chlorobi (■), Chloroflexi (■), Crenarchaeota (■), Cyanobacteria (■), Deinococcus-Thermus (■), Deltaproteobacteria (■), Dictyoglomi (■), Epsilonproteobacteria (■), Euryarchaeota (■), Fibrobacteres (■), Firmicutes (■), Fusobacteria (■), Gammaproteobacteria (■), Nitrospirae (■), Planctomycetes (■), Spirochaetales (■), Thermodesulfobacteriales (■), Thermotogales (■), Verrucomicrobiae (■)



原核生物1,502種的全配列の断片化サイズ5kbについての縮退4連塩基のSOM



Prochlorococcus *Shewanella* *Synechococcus* *Burkholderia*

Venterらの5kb以上配列(4258件; その環境での優先種の配列)のマッピング

Metagenome解析によって得られた環境由来DNA配列による微生物群集の比較

1. サルガッソー海の微生物集団

2. 土壌と深海(鯨骨生物群集)の微生物集団

Tringeらは、土壌・深海由来の試料に関して、異なる環境間での微生物の多様性および相対的存在量の比較解析を行った。



3. モントレー海の微生物集団(TIGR)

TIGRにおいて、モントレー海の異なる深度(0m, 80m, 100m, 750m)より採取した試料に関して、BACクローンの作成を行い、BAC末端の配列決定を行い、比較解析を行っている。

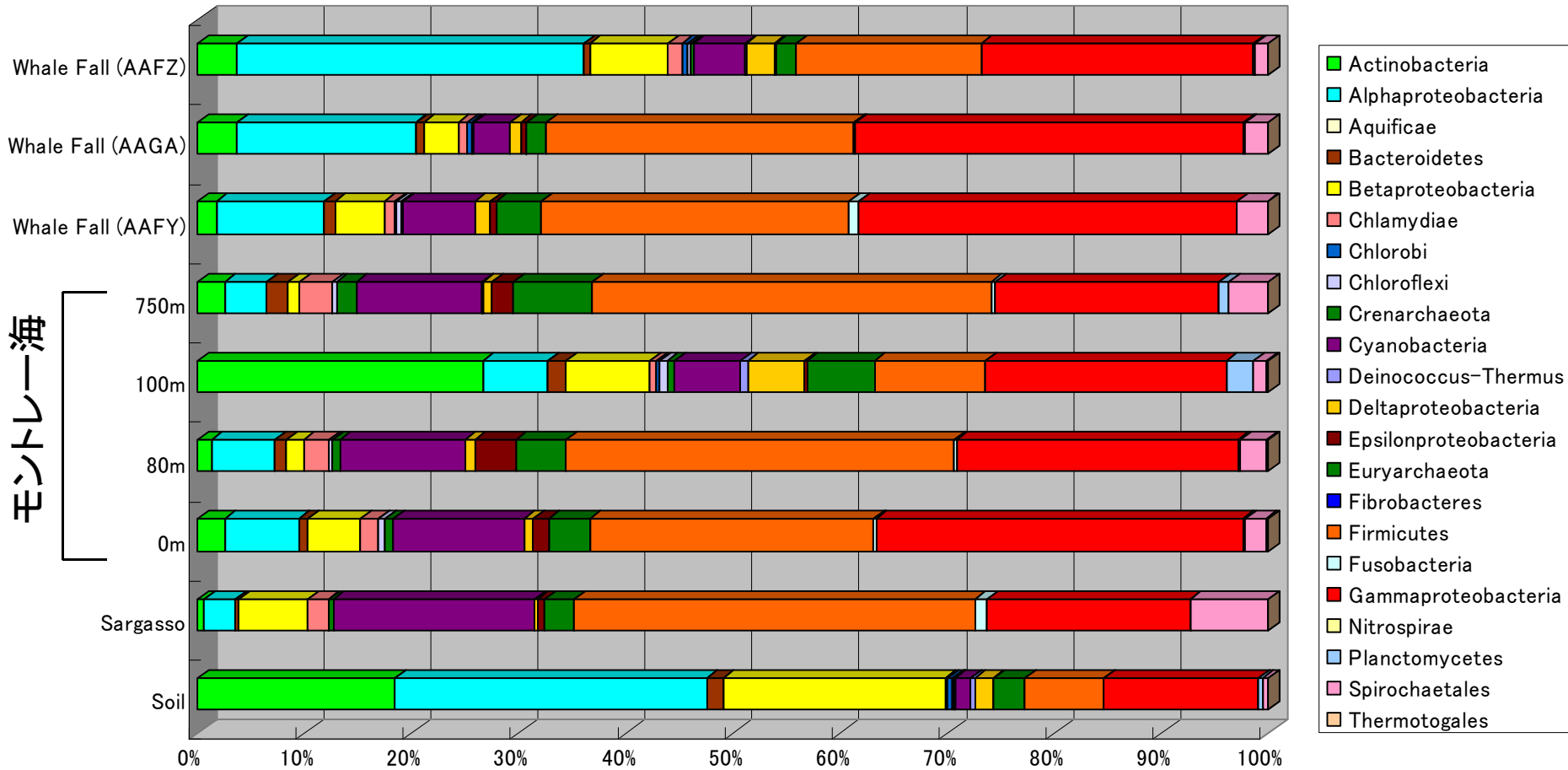
1. Venter et al. (2004).

Environmental Genome Shotgun Sequencing of the Sargasso Sea, Science 304: 66-74

2. Tringe et al (2005).

Comparative Metagenomics of Microbial Communities, Science 308: 554-557

SOMにより予測された系統群数による各環境間での微生物群集の比較



推定された系統群由来とされる配列の割合

評価方法は、SOMマップ上に分類された格子点±1に分類されていた断片配列の系統群の数により系統群の推定を行っている。予測率100%とは、分類された格子点±1に同一の系統群しか分類されていなかった場合であり、今回は予測率50%で算出している。

メタゲノム解析による常在菌や病原菌の多様性の解明を目的にして、SOM解析を行う

- DNA配列の系統群の推定が可能
 - (オーソロガスな遺伝子配列セットを必要としない)
- 多種多様な体内や自然環境中の共生細菌等の多様性の可視化が可能
- 異なる環境間での微生物群集の比較解析が可能

地球上に生息する微生物叢の解明

ヒトの腸内に生息する腸内微生物叢の解明

結論

自然科学の進歩により、原子・分子のミクロのレベルで生命現象が研究できる時代になってきた。そして、この分野の研究は国民生活の向上に直接寄与できる可能性が高まってきた。