予測する生命科学・医療および創薬基盤
「京」を用いた計算生命科学

Computational Life Science using HPC

HPC: High Performance Computing

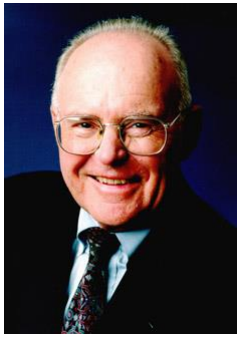
= Science using Supercomputer

Hardware

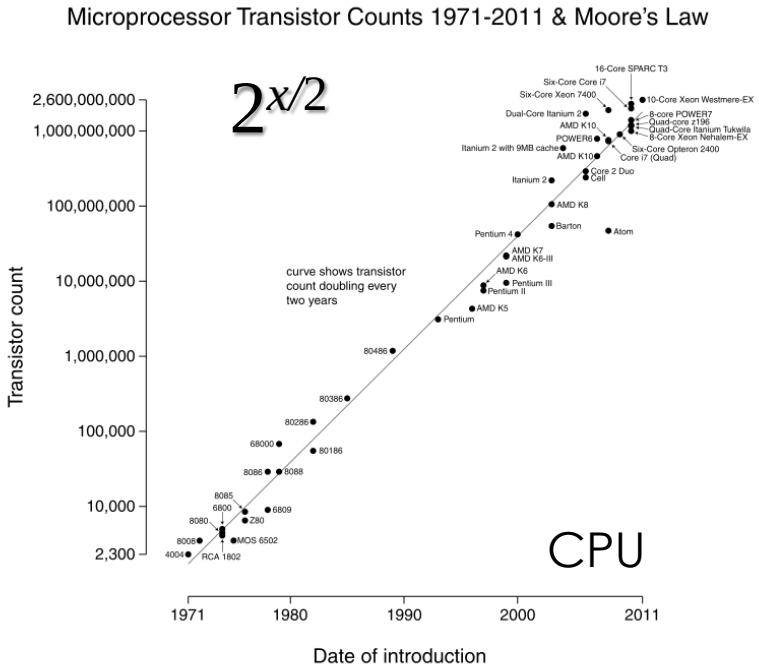
Simulation and Data Analysis

0. Introduction to Supercomputing
1. 計算科学：Capacity Biosupercomputing
2. 生命科学：Connection of Biological Layers

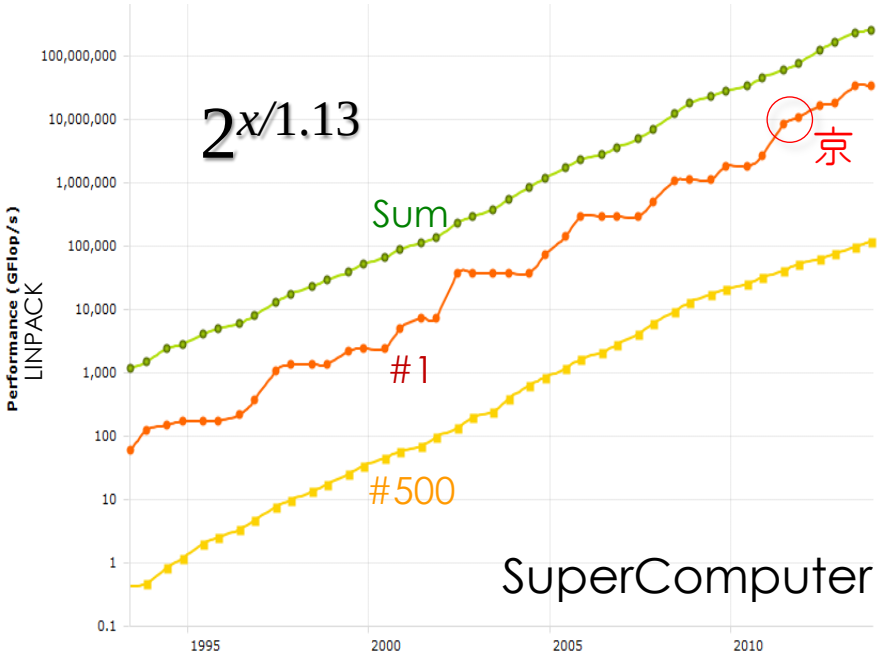
横浜市立大学・理研HPCI
木寺 詔紀



Moore's Law:　発展の推進力（自己目的化？）

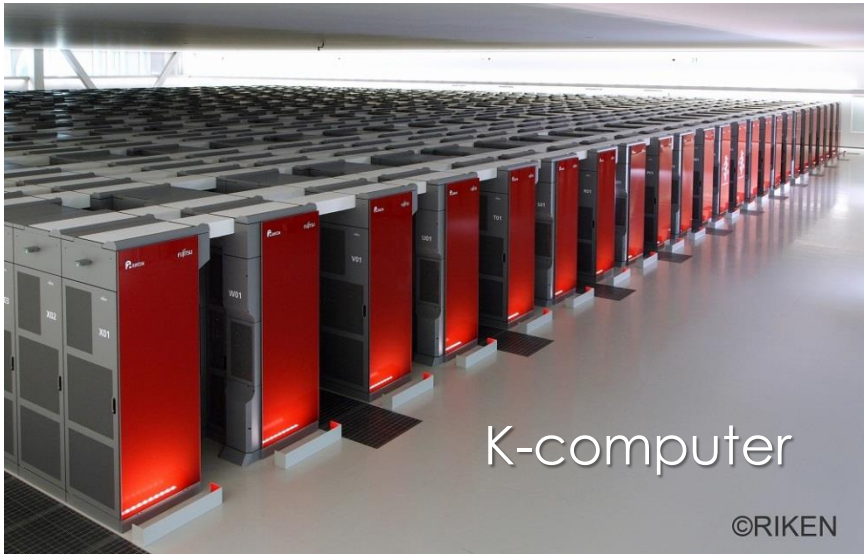


http://en.wikipedia.org/wiki/File:Transistor_Count_and_Moore%27s_Law_-_2011.svg



http://www.top500.org/statistics/perfdevel/#.U4W11yj-stU

HPC in Japan



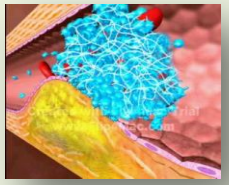
10 PFlops

~3 PFlops

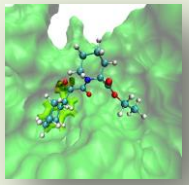
HPCI戦略5分野

1 予測する生命科学・医療および創薬基盤

予測医療と革新的創薬



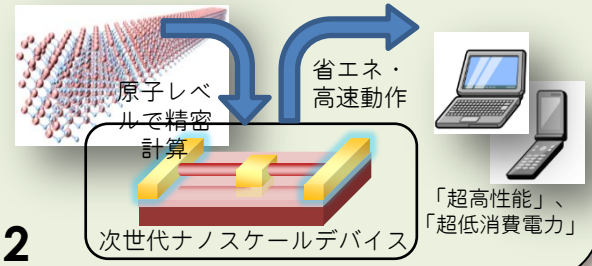
血栓成長による血管閉塞シミュレーション



薬候補のタンパク質への結合シミュレーション

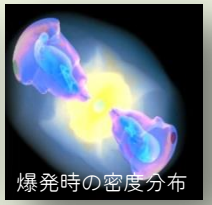
2 新物質・エネルギー創成

世界に先駆けた次世代デバイスを提唱



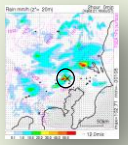
5 物質と宇宙の起源と構造

大質量星の超新星爆発の解明

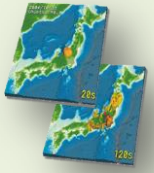


3 防災・減災に資する地球変動予測

集中豪雨や地震の予測



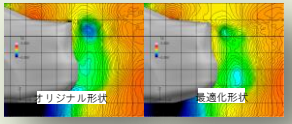
集中豪雨の再現実験



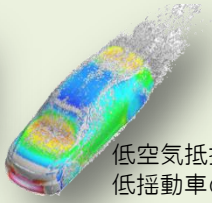
新潟県中越地震による強い揺れの広がり

4 次世代ものづくり

設計プロセスの革新



車体後部周りの超精緻解析による最適形状の究明



低空気抵抗、低揺動車の開発

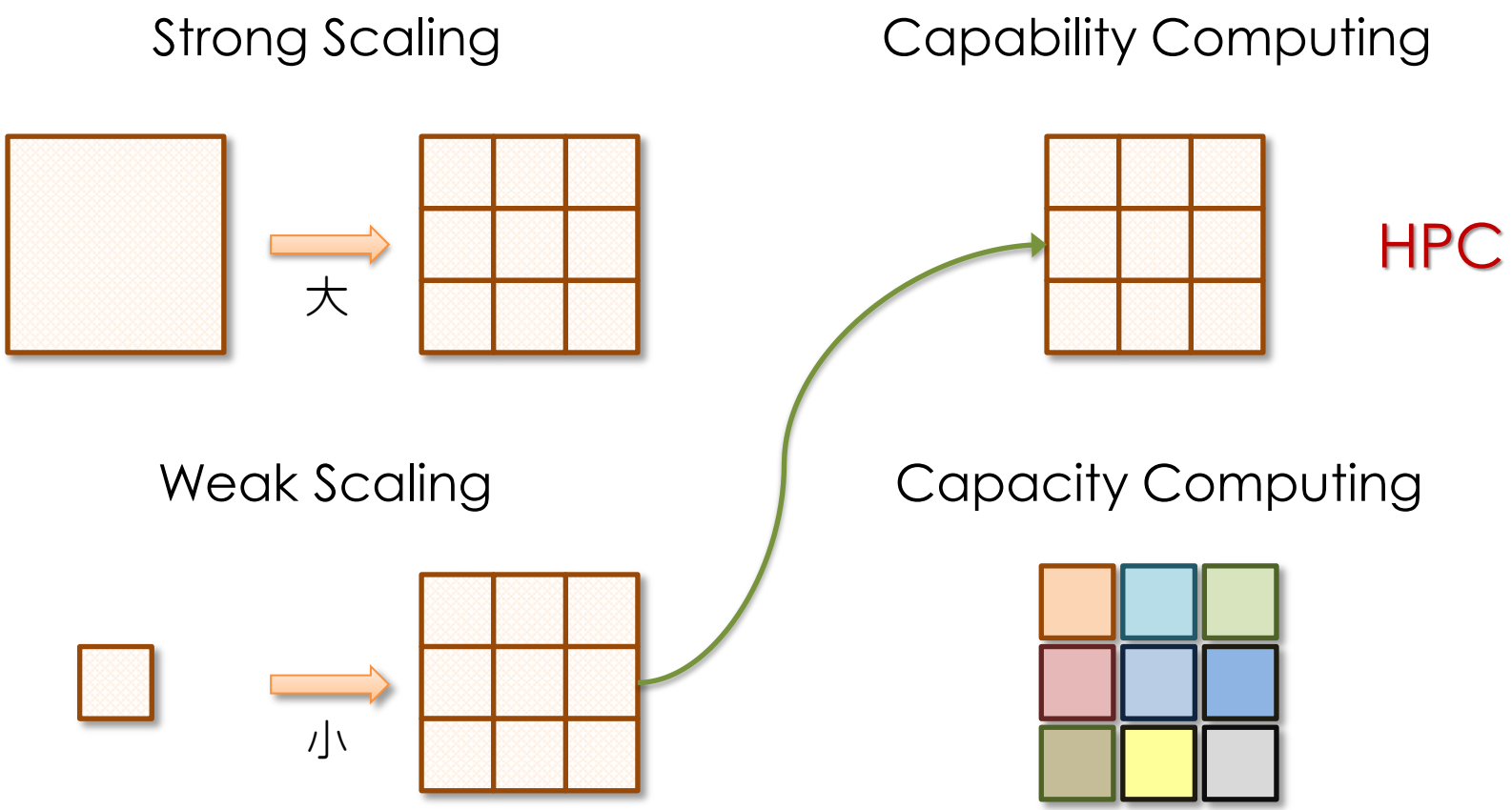
HPC

並進対称
均質
Scalable

↓

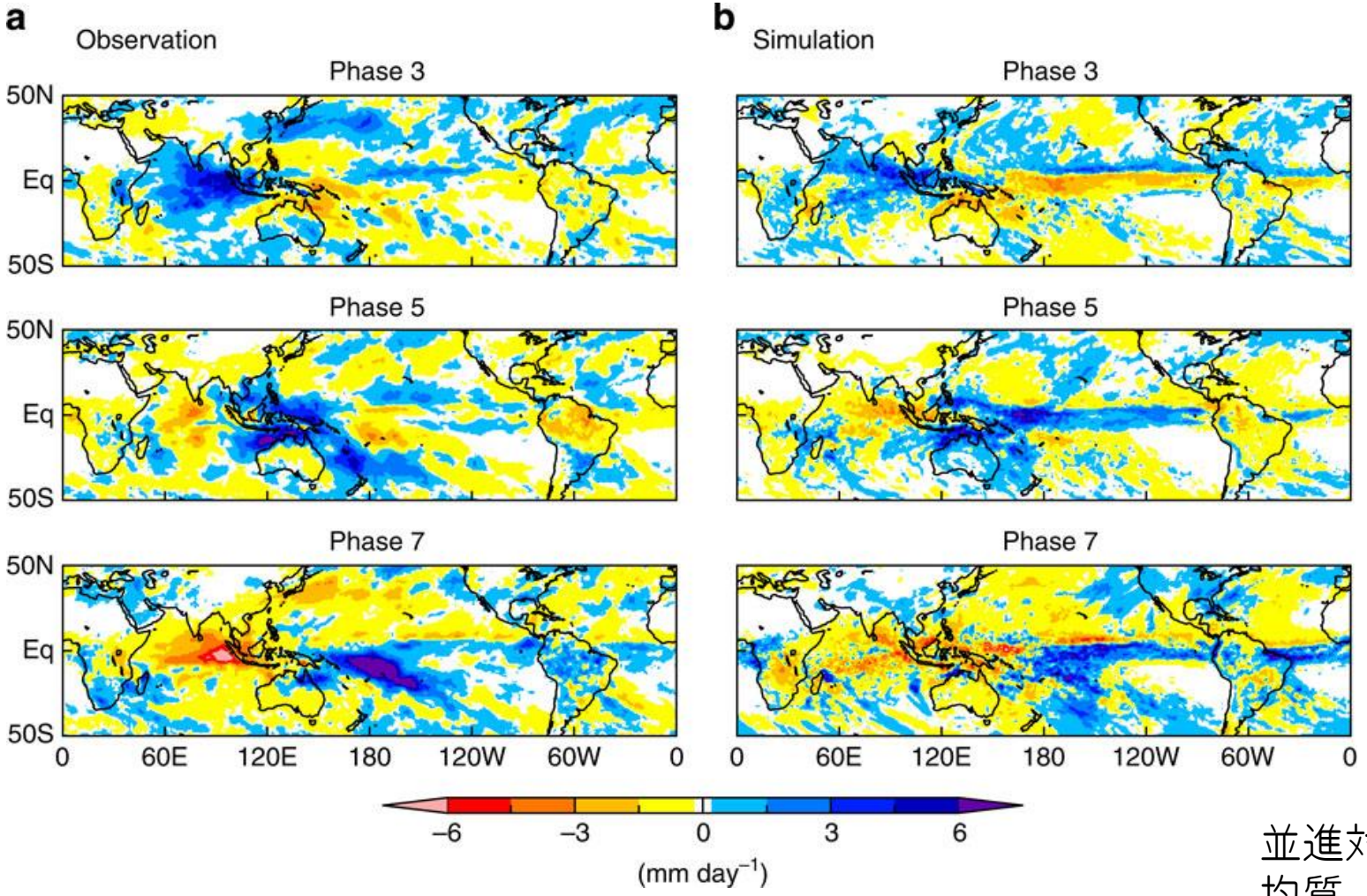
Capability
Computing

Parallel Computing
Data Communication



Capability Computing: Weak Scaling (分野3)

Madden–Julian Oscillation prediction skill of a new-generation global model demonstrated using a supercomputer

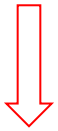


並進対称
均質
Scalable

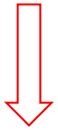
Tomoki Miyakawa, et al., Nature Communications 5, 3769 (2014)

ab initio simulation

確定した間違いのない計算を与えられた量だけ計算すれば、予定された結果が得られる：一度の計算（一通りの境界条件）で問題の本質を明らかにすることができる



全計算機資源の集中 = Capability Computing



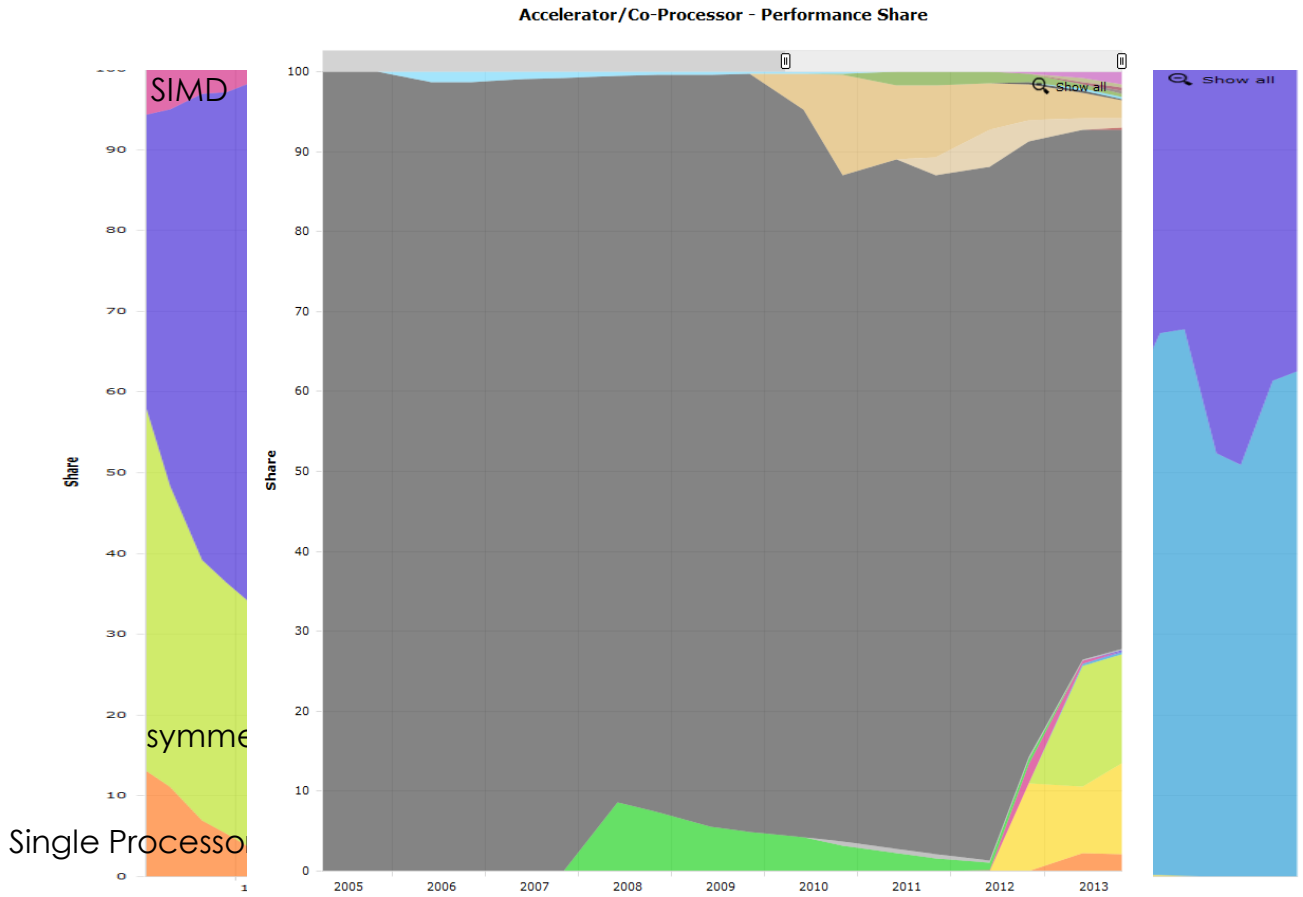
Batch処理 = 最大効率で計算資源を活用する

Moore's Lawはすでに限界？

並列技術（SIMD, … , Accelerator）がそれをカバーしてきた。下方展開は、Clusterが支えてきた。そして、

CPU：汎用

GPU：専用



・Codingの困難さ

・下方展開の困難さ

・Superocmputerが専用と汎用に分岐していく？

・「京」は最後の最大の汎用のsupercomputer ？

http://www.top500.org/statistics/overtime/#.U4W6NSj-stU

1. 計算科学：Capacity Biosupercomputing

HPCI戦略プログラム　戦略分野１

予測する生命科学

生命現象は、多様な要素からなる巨大で複雑なシステム(ゲノム、生体分子、細胞、臓器などの様々な階層)によって担われています。生命システムについての実験情報をもとにしてそのモデルを計算機中に構築し、稼働させることで生命現象をシミュレートすることが我々の研究手段です。シミュレーションで扱うことのできる問題は、用いる計算機の性能によって限定され、部分システムへの分割、モデルの粗視化、短時間現象への限定などが必要となります。スーパーコンピュータ「京」は、そのように問題を限定することなしに、システム丸ごとのシミュレーションを可能とします。「京」によるシミュレーション結果は、実験情報の背後に隠されていた生命現象の動作原理を明らかにし、結果として生命現象の予測、制御の可能性を与えます。特に、生命現象として疾病を対象することで、予測、制御のシミュレーションを、医療、創薬への貢献につなげることが我々の最終的な目的です。

参加機関：理化学研究所、名古屋大学、東京大学、日本原子力研究開発機構、沖縄科学技術研究基盤整備機構、東京工業大学、産業技術総合研究所、大阪大学、国立遺伝学研究所、横浜市立大学、京都大学、統計数理研究所（順不同）

HPCI戦略プログラムは、スーパーコンピュータ「京」を中心としたHPCI(High Performance Computing Infrastructure)を最大限に活用することによって、戦略的に取り組むべき5つの研究分野において画期的な成果を産み出し、計算科学技術の飛躍的な発展を目指す文部科学省のプログラムです。

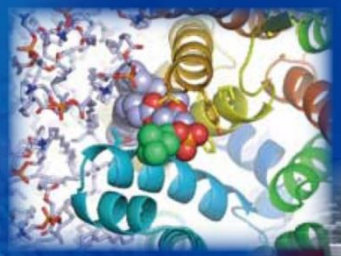
ライフサイエンス分野は、理化学研究所を代表機関として、全国の大学や研究機関が共同して研究を行います。

さらに、計算科学技術コミュニティの発展を目指して、技術講習会やネットワーク形成・拡大にも取り組みます。

1.研究開発課題

1 細胞内分子ダイナミクスのシミュレーション

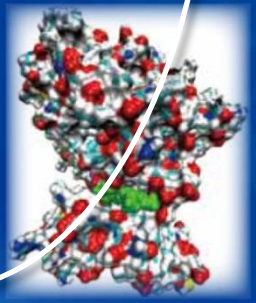
マルチスケール分子ダイナミクス・シミュレーションと一分子粒度シミュレーションを解析ツールとして、細胞内環境下における生体分子の構造と機能を研究します。生体膜を介した物質輸送、タンパク質／DNA相互作用、シグナル伝達が、代表的な研究ターゲットです。細胞機能の理解や薬剤設計等への寄与が期待されます。

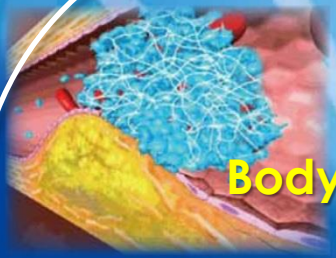


Molecular

2 創薬応用シミュレーション

新規化合物設計方法の構築、化合物とタンパク質の結合部位同定法の確立を行い、複数の標的タンパク質について薬候補化合物の設計・検証を行う。また、柔らかい標的部位を持つタンパク質に対するリガンド結合予測法を開発する。効率の良い薬剤設計手法の確立が期待されます。

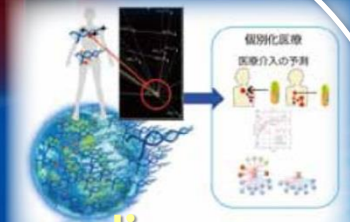




Body/Organ

3 予測医療に向けた階層統合シミュレーション

全身筋骨格ー脳神経シミュレータ、血栓シミュレータ、心臓シミュレータ、全身血管網シミュレータを結合し、複雑な生命現象のより深い理解と幅広い疾患に対応できる統合シミュレータを構築します。狭心症、心筋梗塞等の心臓疾患や、神経疾患に伴う運動機能障害等の治療法の検討や薬効の評価、予後の予測への貢献が期待されます。



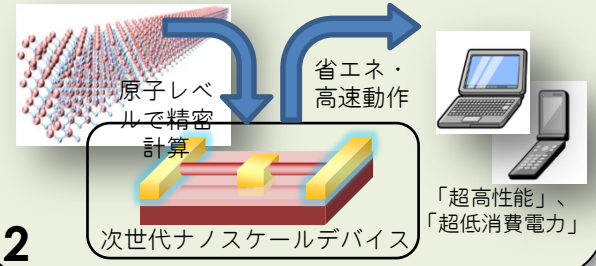
Bioinformatics

4 大規模生命データ解析

次世代シークエンサーなどの先端機器を利用して得られるデータを大規模に解析するための手法を開発し、病気や生物多様性の理解、細胞やがんの個性の描出、生体分子ネットワークの理解を目指します。個別化医療やゲノム情報の産業利用への貢献が期待されます。

新物質・エネルギー創成

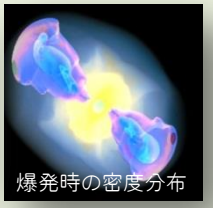
世界に先駆けた次世代デバイスを提唱



2

物質と宇宙の起源と構造

大質量星の超新星爆発の解明

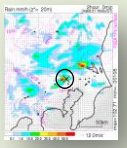


5

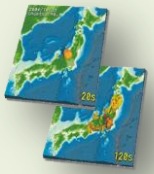
防災・減災に資する地球変動予測

3

集中豪雨や地震の予測



集中豪雨の再現実験

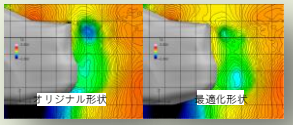


新潟県中越地震による強い揺れの広がり

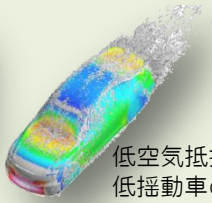
次世代ものづくり

4

設計プロセスの革新



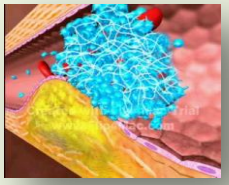
車体後部周りの超精緻解析による最適形状の究明



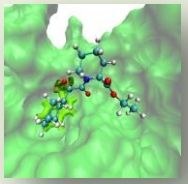
低空気抵抗、低揺動車の開発

予測する生命科学・ 医療および創薬基盤

予測医療と革新的創薬



血栓成長による血管閉塞シミュレーション



薬候補のタンパク質への結合シミュレーション

1

並進対称
均質
Scalable

並進非対称
不均質
Unscalable

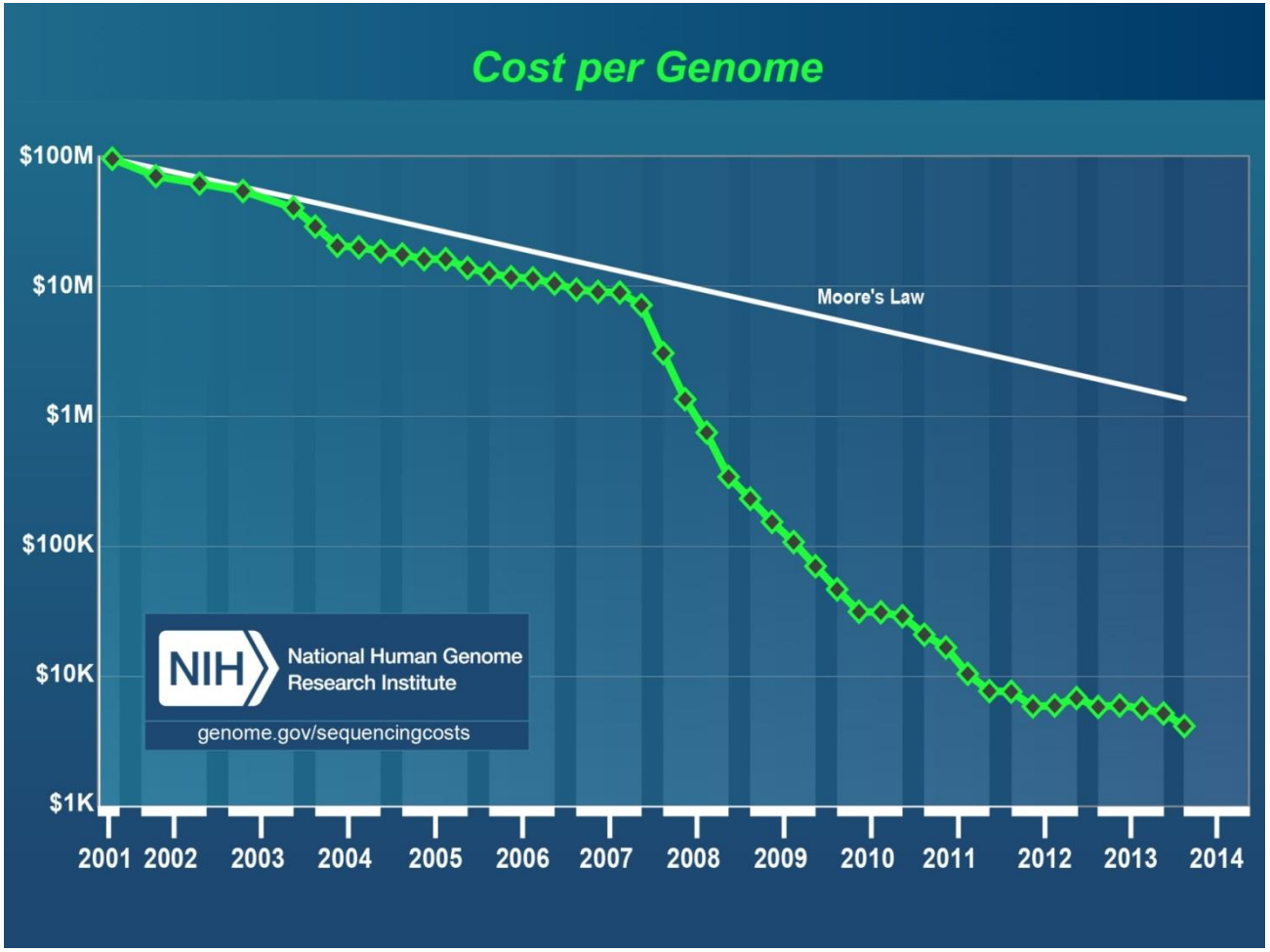
意味のあるweak scalingは困難

生命科学は積極的にCapacity Computing

- **複雑系**　多数のレプリカの計算
  ensemble simulation, landscape picture（data assimilation）

- **非平衡性**　path ensemble (多数のレプリカの計算)
  システムは外部からの情報、物質の伝達によって機能発現する。
  孤立系ではなく外部からの摂動を受ける非平衡系。
  緩和過程のensembleは1度の長時間simulationでは達成できない。

Big Data in Life Science



http://www.genome.gov/

生命科学は積極的にCapacity Computing

- **複雑系**　多数のレプリカの計算
  ensemble simulation, landscape picture (in a parameter space)

- **非平衡性**　Path ensemble (多数のレプリカの計算)
  システムは外部からの情報、物質の伝達によって機能発現する。
  孤立系ではなく外部からの摂動を受ける非平衡系。
  緩和過程のensembleは1度の長時間simulationでは達成できない。

- **枚挙の論理**　類似した計算を繰り返す
  Dataごとの（Modelごとの）計算結果
  - Variation of genes
  - Homologous proteins
  - Epigenetic difference
  - Change in cellular environments
  - Normal system / Disease related system
  - Patients A, B,…

  医療・創薬 ＝ 個別化

- **変化、差異**　類似した計算を繰り返す
  変化、差異、異常-正常の違い（Δ、ΔΔが対象である）

Data-driven simulation
vs. *ab initio* simulation

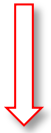
計算の環境

与えられたDataごとの計算　　Adjoining data production

Dataに特異的な結果

計算の手法

多数のレプリカを用いた計算　　Capacity Computing



Interactive処理
On-demand
HPC Cloud
Data I/O (Machine, Internet)

最大効率で結果を生み出す

HPCはこの要請にこたえるべきである

Capacity Biosupercomputing

大規模生命情報解析　宮野悟（東大）

遺伝子ネットワーク推定

Public Gene Expression Data

3,470,000 node·hours （K Computer）

690 networks

がん細胞の薬剤感受性の遺伝子ネットワーク解析

Sanger Gemonics of Drug Sensitivity in Cancer
(13,435 genes x 728 cell lines x 142 drugs)

2,020,000 node·hours （K Computer）

Profiles for 101 drugs

がん細胞における融合遺伝子の網羅的検出

Cancer Cell Line Encyclopedia (Broad Institute) 780 samples

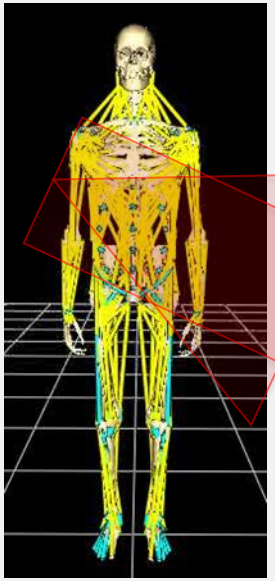
976,102 node·hours （K Computer）

12,600 fusion genes

計算生命科学：階層の接続

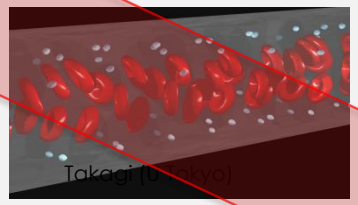
data-based model

全身



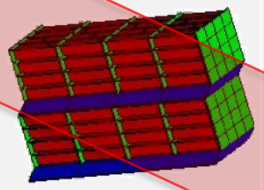
Nakamura (U Tokyo)

臓器



Takagi (U Tokyo)

細胞

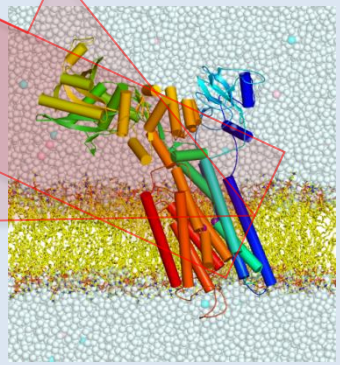


Hisada (U Tokyo)

実体モデルのシミュレーションによって生命科学の目的である生命階層の接続がどのように可能か

ab initio model

分子



Sugita (Riken)

計算機顕微鏡 = 因果律（巨視的→微視的）

計算機実験 = 予測（微視的→巨視的）

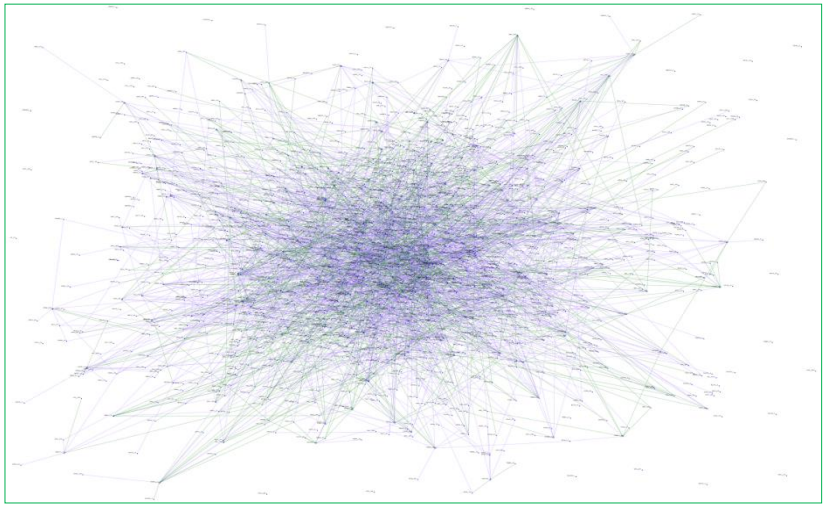
生命情報科学

CGGGCTTTTTTTTTCGACCAAAGGTAACGAGGTAACAACC
ATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATG
CAGAACGTTTTCTGCGTGTTGCCGATATTCTGGAAAGCAA
TGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCC
GCCAAAATCACCAACCACCTGGTGGCGATGATTGAAAAAA
CCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGC
CGAACGTATTTTTGCCGAACTTTTGACGGGACTCGCCGCC
GCCCAGCCGGGGTTCCCGCTGGCGCAATTGAAAACTTTCG
TCGATCAGGAATTTGCCCAAATAAAACATGCGGGCTTTTT
TTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGT
TGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTT
TCTGCGTGTTGCCGATATTCTGGAAAGCAATGCCAGGCAG
GGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCA
CCAACCACCTGGTGGCGATGATTGAAAAAACCATTAGCGG
CCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATT
TTTGCCGAACTTTTGACGGGACTCGCCGCCGCCCAGCCGG

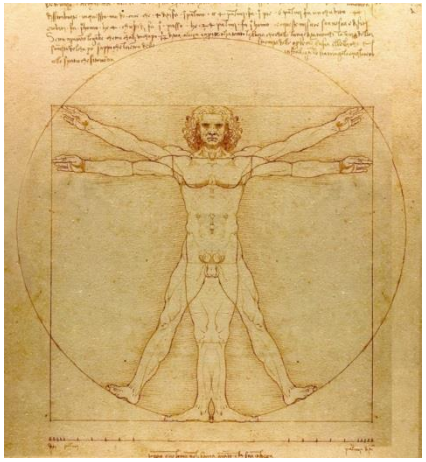
Genotype
（微視的）

Systems
（model）

Phenotype
（巨視的）

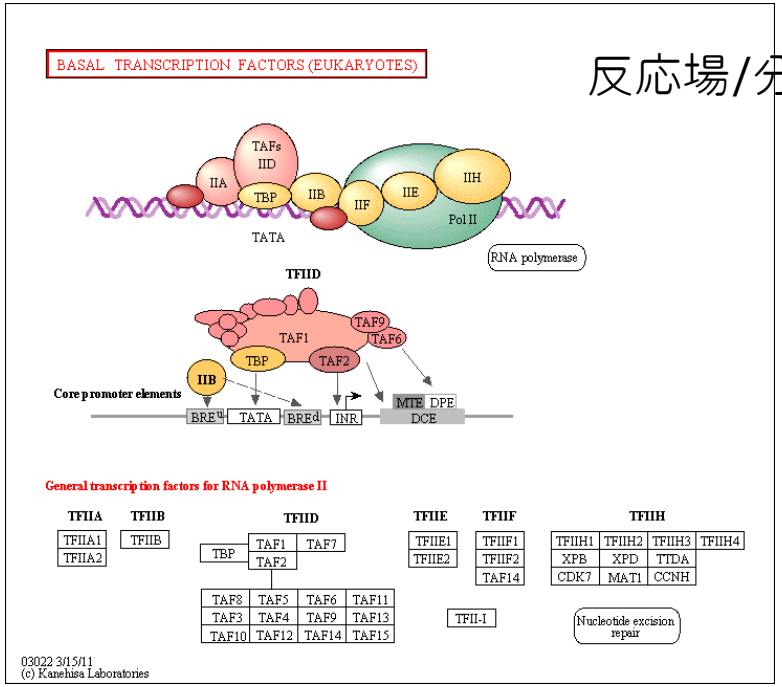


遺伝子ネットワーク/宮野悟

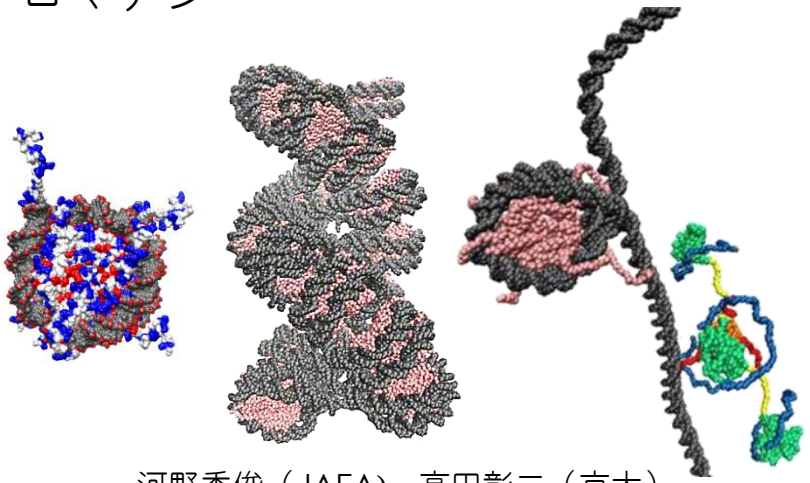


システムの実体化
分子シミュレーション

階層接続：システムの実体化としての分子シミュレーション

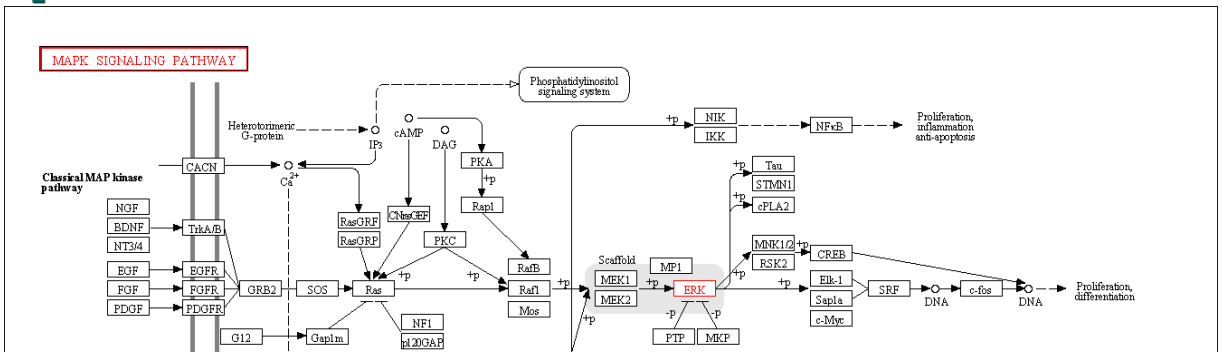


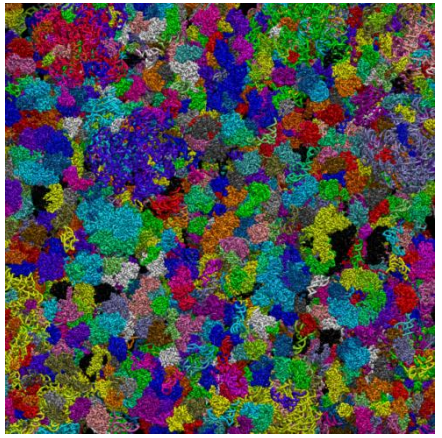
反応場/分子機械：クロマチン



河野秀俊（JAEA)、高田彰二（京大）

信号伝達/拡散：分子混雑





杉田有治（理研)

階層接続：マルチスケール・マルチフィジックス シミュレーション　ー　UT-Heart（久田 (東大)）

分子、細胞、臓器それぞれのスケールでの原因による疾患をシミュレートする

molecule

organ

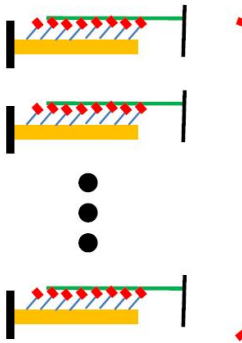
確率的なモデル（統計力

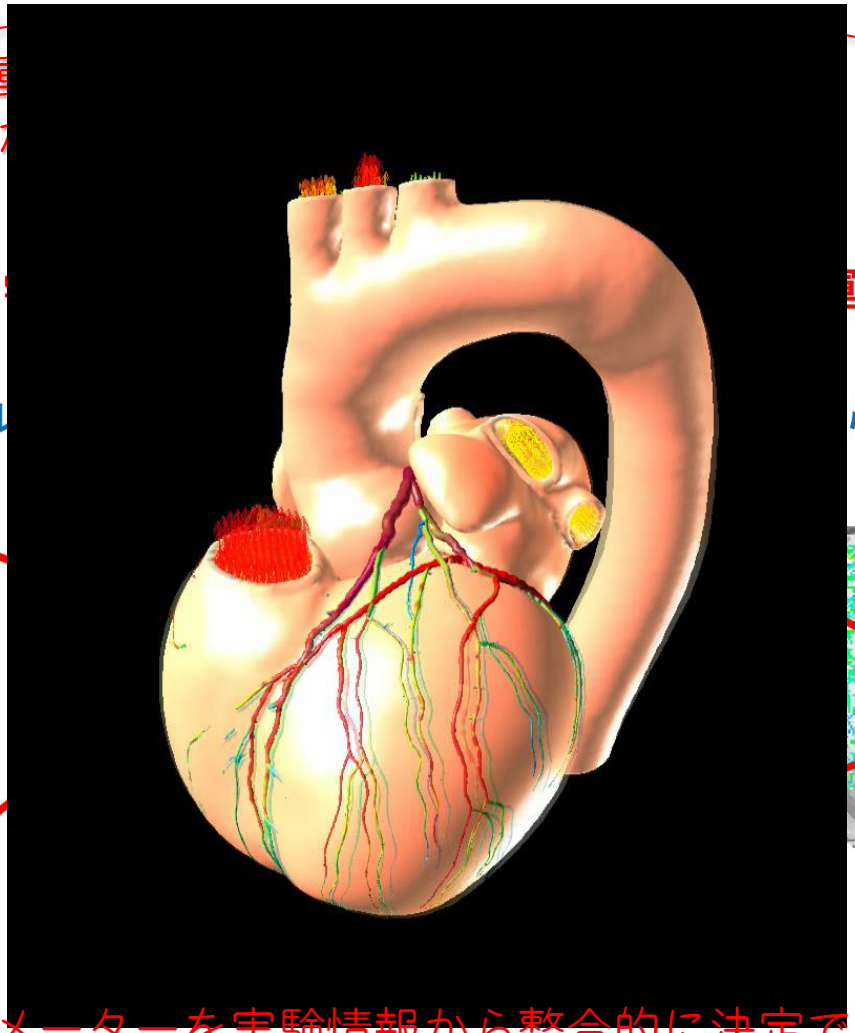
運動方程式）

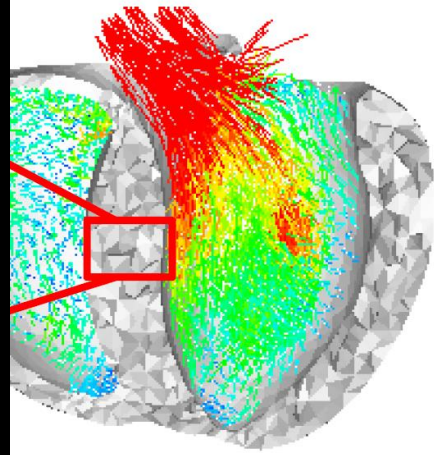
モンテカルロ法（MC）

32個のサルコメアモデル
を筋原線維要素に

心室壁の相互作用を解析





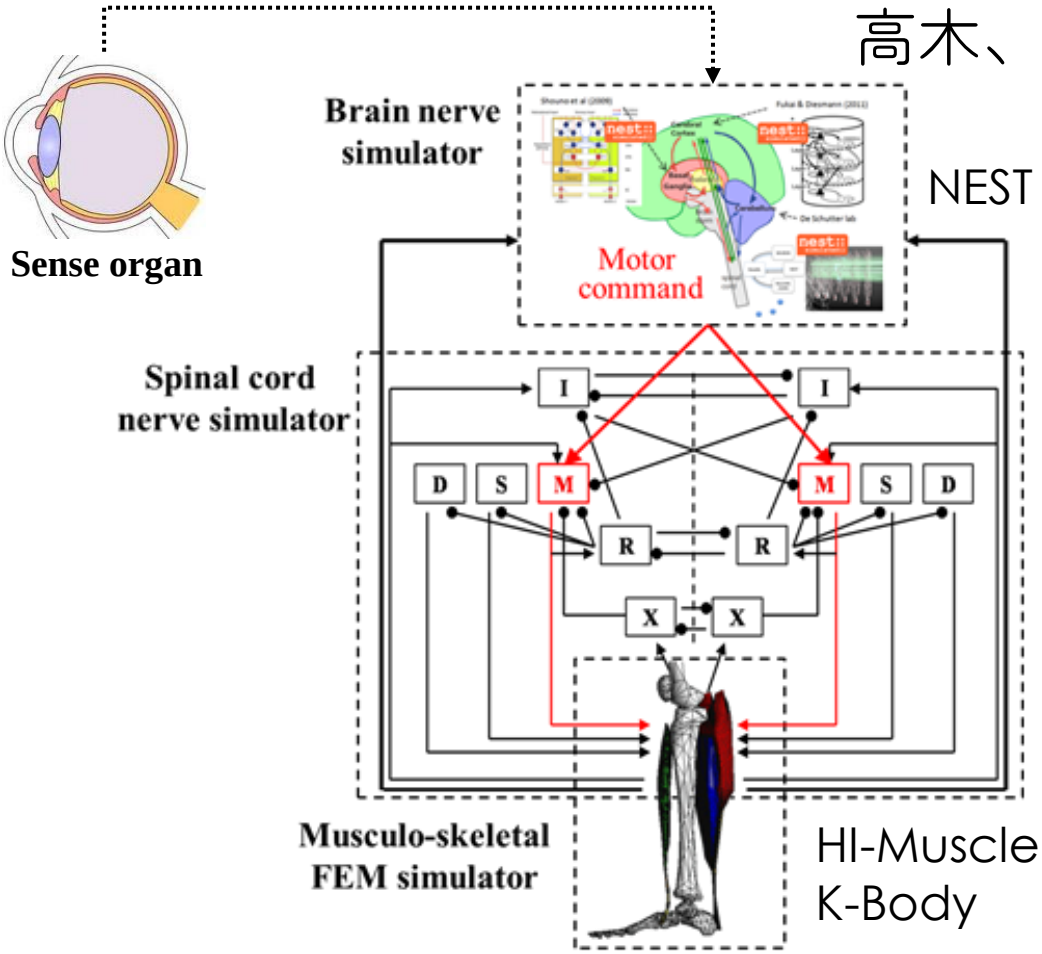


パラメーターを実験情報から整合的に決定できる

© 鷲尾，久田（東大）

階層接続：マルチスケール・マルチフィジックス シミュレーション　ー 全身統合モデル

高木、中村 (東大)、銅谷 (OIST)



異なった原理で作動するシミュレーターの統合

Hodgkin-Huxley model/integrate and fire model：ion channels

神経回路

脊髄反射

運動器官

脳・神経系の異常に伴う運動障害

Parkinson's disease

各シミュレータの時間ギャップが少ない

パラメーターを実験情報から整合的に決定できる

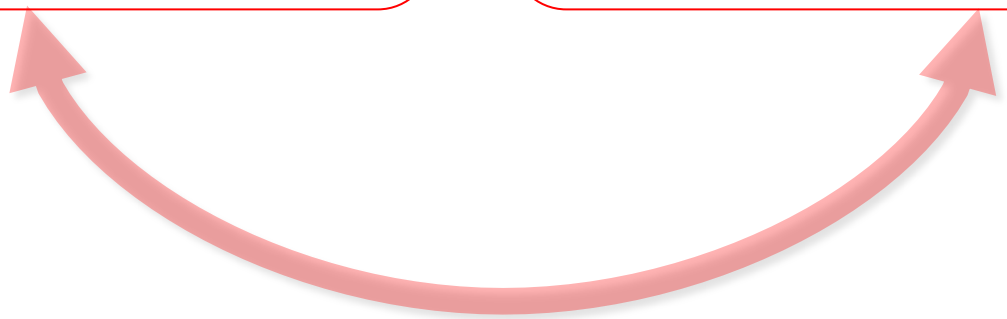
Data-driven simulation

HPC Simulations

「情報に基づいたシミュレーション」
data-based model

Life Science Data Analysis

「シミュレーション技術を用いた情報解析」
dynamic system

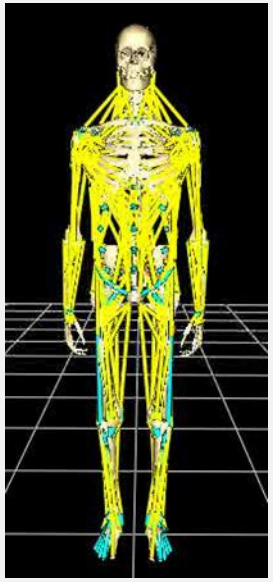


計算生命科学：階層の接続

Data-driven simulation

ab initio simulation

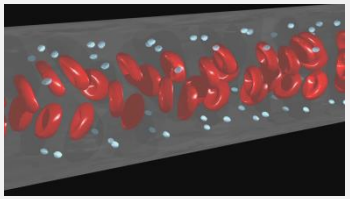
全身



Nakamura (U Tokyo)

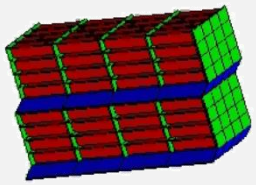
data-based model

臓器



Takagi (U Tokyo)

細胞



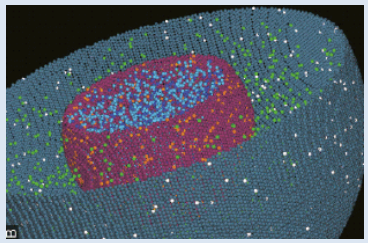
Hisada (U Tokyo)

分子

構造の遷移状態モデル
電流の微分方程式

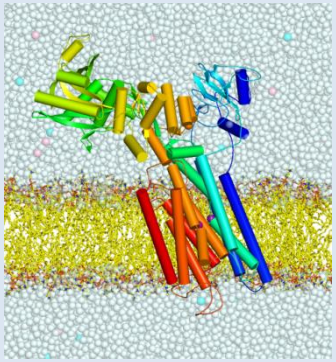
ab initio model

細胞



Takahashi (Riken)

分子



Sugita (Riken)

個別化

予測する生命科学・医療および創薬基盤

SCLS SUPERCOMPUTATIONAL LIFE SCIENCE

課題２　創薬応用シミュレーション

代表：藤谷 秀章　東京大学 先端科学技術研究センター

藤谷 秀章
（東京大学 先端科学技術研究センター）
「京」での薬候補の評価　高精度自由エネルギー計算

理化学研究所
創薬・医療技術基盤プログラム
疾患標的探索、化学・細胞アッセイ、化合物設計など

課題１ 細胞内分子ダイナミクスのシミュレーション

代表：杉田 有治　理化学研究所 杉田理論分子科学研究室

1) 細胞環境を考慮した信号伝達経路のモデリング

杉田 有治
（理化学研究所 杉田理論分子科学研究室）
細胞内分子混雑

高橋 恒一
（理化学研究所 生命システム研究センター）
信号伝達経路全体モデリング

高田 彰二
（京都大学大学院 理学研究科）
リン酸化複合体

林 重彦
（京都大学大学院 理学研究科）
反応自由エネルギー解析

2) 核内混み合い環境でのヌクレオソーム、クロマチンの機能発現機構

河野 秀俊
（日本原子力研究開発機構）
ヌクレオソーム

高田 彰二
（京都大学大学院理学研究科）
クロマチン

池口 満徳
（横浜市立大学生命医科学研究科）
核内DNA結合蛋白質

課題３　予測医療に向けた階層統合シミュレーション

代表：高木 周　東京大学大学院工学系研究科

1) 心筋梗塞・脳梗塞のマルチスケールシミュレーション

後藤 信哉
（東海大学医学部内科学系）
血栓症のモデリング

高木 周
（東京大学大学院工学系研究科）
マルチスケール血栓シミュレータ

2) 心疾患の治療法・薬効評価のためのマルチスケール・マルチフィジックス心臓シミュレーション

久田 俊明
（東京大学大学院新領域創成科学研究科）
心臓シミュレーションの機能拡張

3) 神経疾患による運動機能障害解明のための全身筋骨格 - 神経系統合シミュレーション

中村 仁彦
（東京大学大学院情報理工学系研究科）
全身筋骨格神経系シミュレータ

野村 泰伸
（大阪大学大学院基礎工学研究科）
パーキンソン病のモデリング

銅谷 賢治
（沖縄科学技術大学院大学 神経計算ユニット）
大脳皮質・大脳基底核・小脳の神経回路のモデル化

高木 周
（東京大学大学院工学系研究科）
筋繊維からのマルチスケール骨格筋シミュレータ

課題４　大規模生命データ解析

代表：宮野 悟　東京大学 医科学研究所

1) 大規模データ解析によるがんのシステム異常の網羅的解析とその応用

宮野 悟
（東京大学 医科学研究所）

2) 大規模生体分子ネットワーク解析による脂肪細胞組織の刺激応答の網羅的解析とその応用

松田 秀雄
（大阪大学大学院情報科学研究科）

3) 次世代シーケンサーデータ解析のための情報処理システムの開発

秋山 泰
（東京工業大学大学院情報理工学研究科）