

# 次世代生命体統合シミュレーションソフトウェア の研究開発

理化学研究所  
次世代計算科学研究開発プログラム



1

## そもそもの動機

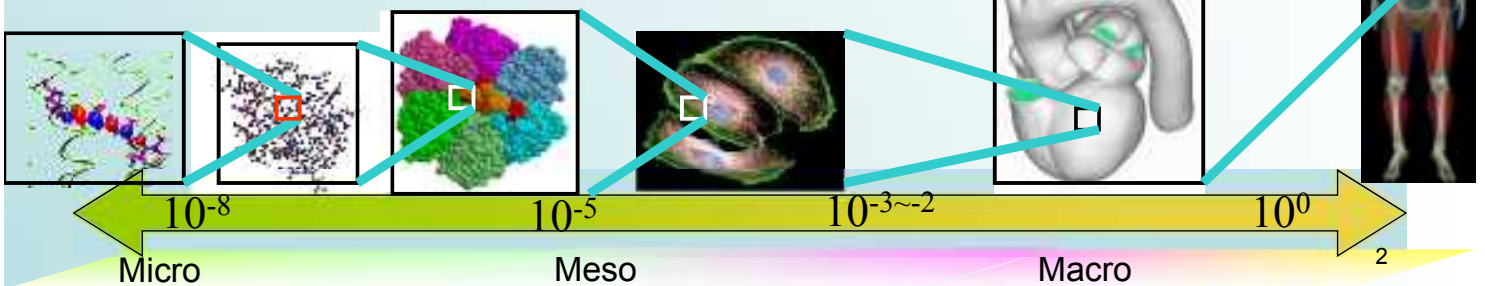
生命現象は最も複雑で難しい問題

複雑で美しい振舞いを示す**超**多体系多階層問題



スーパーコンピュータを使って、この複雑な生命現象を解析

記述する生物学から、予測する生物学へ



2

## 研究開発の概要と達成目標

基礎方程式に基づく解析的アプローチと、大量の実験データから未知の法則に迫る実験データから解析へのアプローチ、さらには多階層を連結するアプローチにより、異なるスケールの研究と実験データを統合的かつ有機的に結びつけ、ペタスケールという桁違いの性能を持つスーパーコンピュータの性能をフルに発揮し、生体で起こる種々の現象を理解し医療に貢献するためのソフトウェアを開発する。



3

## これまでの開発の概要

- 経緯
  - 2006年10月から開発を開始、この時、分子・細胞・臓器全身・データ解析融合の4チーム。年度内に高度化チームを追加し、ソフト開発を支援
  - 2007年10月の脳神経系チームを追加
  - 2010年度中間評価
  - 2011年4月から京を使った開発が開始
- マネージメント
  - ソフトウェア開発の進捗状況を的確に把握し、支援
  - ソフトウェアにプライオリティーを設定、重点化

4

# 各チームで取り組む課題 (2008年での整理)

## 整理前

計算手法

分子スケール	QM/MM
	全電子量子化学計算
	ハイブリッドQM/MM法
	分子動力学計算法
	粗視化シミュレーション
	蛋白質の構造サンプリング技術

共通的な開発対象の設定

分子スケール	QM/MM
	全電子量子化学計算
	膜タンパク質・代謝酵素
	分子動力学計算法
	粗視化シミュレーション
	蛋白質の構造サンプリング技術

## 整理後

短期目標

長期目標

計算対象

細胞スケール	細胞シミュレーション統合プラットフォーム
	肝細胞(代謝)
	膵臓β細胞
	イオンチャネル
	血小板

開発対象の重点化による整理見直し

細胞スケール	細胞シミュレーション統合プラットフォーム
	肝細胞・肝小葉(代謝)

長期目標に向けた課題

計算対象

臓器全身スケール	血流シミュレーション
	心臓シミュレーション
	全身ホクセルデータ・全身力学モデル
	超音波伝搬シミュレーション
	肺シミュレーション

臓器全身スケール	循環器系シミュレーション (血球・血小板・血流・血管網・心臓)
	全身ホクセルデータ・全身力学モデル
	超音波伝搬シミュレーション

プログラム

データ解析融合	大規模遺伝子ネットワーク探索ソフトウェア
	生命体データ同化プログラム
	タンパク質ドッキング解析プログラム
	大規模SNP解析ソフトウェア

共通的な開発対象の設定

データ解析	大規模遺伝子ネットワーク探索ソフトウェア
	生命体データ同化プログラム
	肺がんと薬
	タンパク質ドッキング解析プログラム
	大規模SNP解析ソフトウェア

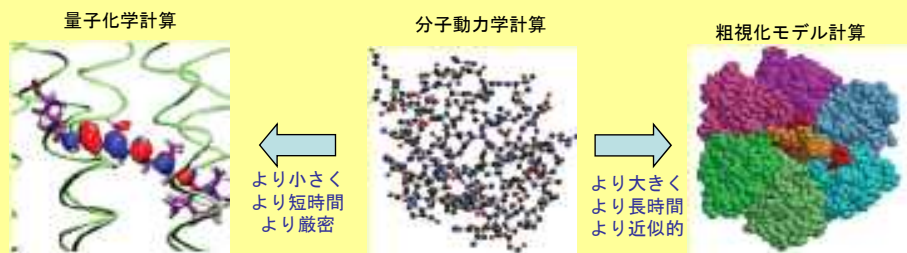
他チームとの連携による効率化

# 2008年度に設定したチーム間連携課題

受提供	データ解析融合	分子スケール	細胞スケール	臓器全身スケール	脳神経系
データ解析融合	肺がんと薬	データ同化技術	生体分子ネットワークモデル	画像処理パラメータ推定	パラメータ推定(モデル化)
分子スケール	タンパク質-リガンド	代謝・膜	膜での輸送モデル	血栓形成	神経可塑性・変性・イオンチャネル
細胞スケール	生体分子ネットワークデータ	代謝反応・膜での物質輸送データ	肝細胞	複数細胞から臓器へ・血栓形成	細胞プラットフォーム供給
臓器全身スケール	医療画像データ	データ提供・ドレッジデリバリ	血流・分泌物輸送	循環器系	細胞の変形
脳神経系	神経細胞データ・局所回路データ	可塑性や神経変性に関するデータ	成長・変形モデル	運動制御臓器制御	細胞から高次機能までの一貫
高度化	分析・高速化	分析・高速化	分析・高速化	分析・高速化	分析・高速化

**開発目標:** 量子化学計算(QM)・分子動力学計算(MM)・粗視化モデル計算(CG)を総合化する技術を開発、タンパク質や細胞の機能発現過程のシミュレーションにつなげ、細胞スケールとの有機的な連携を図る。ターゲット: 多剤排出トランスポータ、脂肪酸代謝酵素反応

**開発計画:** QM、MM、CGの各スケールでプログラムの開発を進めるとともに、それらの手法を結合したQM/MM、MM/CG法によってマルチスケールシミュレーションを実現するためのプログラムを開発する。



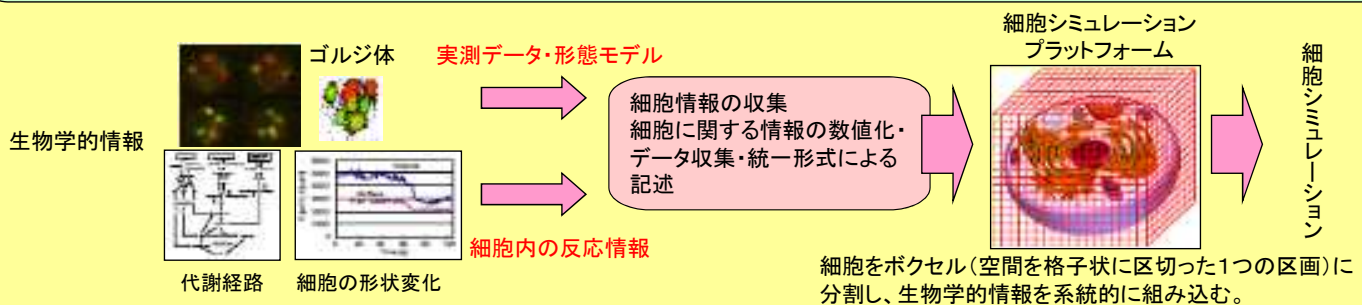
**進捗状況:** QM/MM/CGの3階層それぞれで計算方法をプログラム化し、テストを続けながら開発、8000並列程度まで性能が出るまでになっている。

2011.1.23時点

7

**開発目標:** ペタフロップス級の計算力を活かして細胞を100万のボクセルに区画、そこに包括的な実証データを組み込んだ肝細胞・肝小葉のシミュレーションの実現を目指す。薬剤等の作用や副作用の予測を行うシミュレーションの開発につなげる。

**開発計画:** 生物の空間的情報を入れ込むことが出来るプラットフォームの開発と平行して、プラットフォームに実装するためのデータ収集とモデルの構築を進める。また、プラットフォームと各種データ・モデルとの有機的連携を図る。



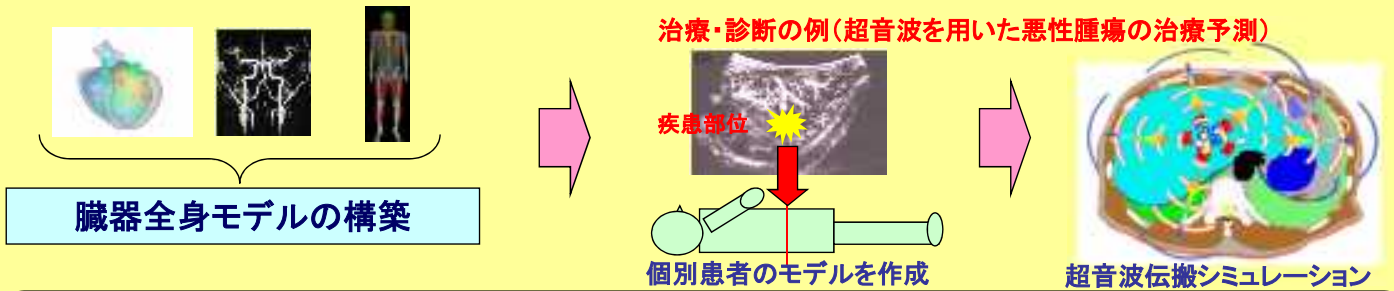
**進捗状況:** 細胞を約3万個のボクセルに分割、複数のオルガネラを表現し、移流拡散方程式による物質移動をシミュレーションできる細胞シミュレーションプラットフォームを開発、E-CELL3をベースにした代謝反応がオルガネラ間も含めてシミュレーションできるようになった。現在、1024並列までの性能テストを終了

2011.1.23時点

8

**開発目標:** ペタフロップス級の計算量を用いて、数時間で実行可能な臓器全身シミュレーションの構築を目指す。これにより、病態予測や治療予測等の医療支援ツールの開発につなげる。

**開発計画:** 血管網、各種臓器、全身を3次元的に再現した臓器全身モデルを構築し、病態予測や治療予測を行う臓器全身モデルを次世代スパコンに実装、循環器系疾患を始めとする種々の病態のシミュレーションと、ガンなどの治療支援シミュレーション(放射線・集束超音波など)につなげる。

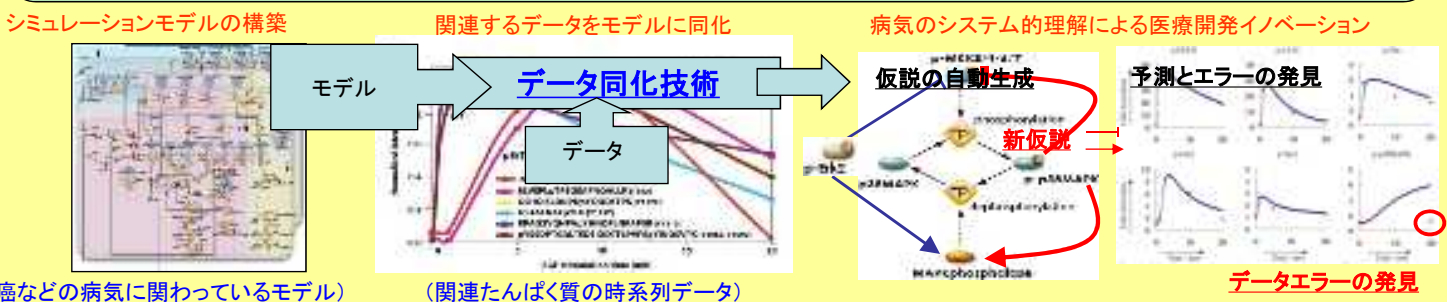


**進捗状況:** 1mm分解能の全身ボクセルモデルを開発すると同時に新しい構造流体連成手法と多媒質内の超音波伝搬解析手法を開発した。これらにより、外科手術シミュレーションや収束超音波・重イオンビームによるガンの治療等へつなげる基盤を構築した。

2011.1.23時点

**達成目標:** ペタフロップス級の計算によって、創薬ターゲット探索や個人差を考慮した医療のための基盤情報技術の構築を目指す。これにより、ヒト全遺伝子を対象とした創薬ターゲット遺伝子探索の実現につなげる。

**開発計画:** 遺伝子ネットワーク推定・タンパク質ネットワーク推定・ゲノム多型解析技術等の研究、ならびにデータ同化技術の開発を進め、開発した解析法やデータ同化技術を次世代スパコンに実装し、データ解析とシミュレーションを融合した技術を開発する。



(癌などの病気に関わっているモデル)

(関連たんぱく質の時系列データ)

データエラーの発見

**進捗状況:** 分子のネットワークを地図として抽出するための方式として、大規模遺伝子ネットワーク探索及びタンパク質構造に基づく相互作用予測の研究を行い、新たな技術開発の成功と新たな並列プログラムの開発により、これらのプログラムが8000を超えるcoreで稼動するようになった。

2011.1.23時点



# 脳神経系研究開発

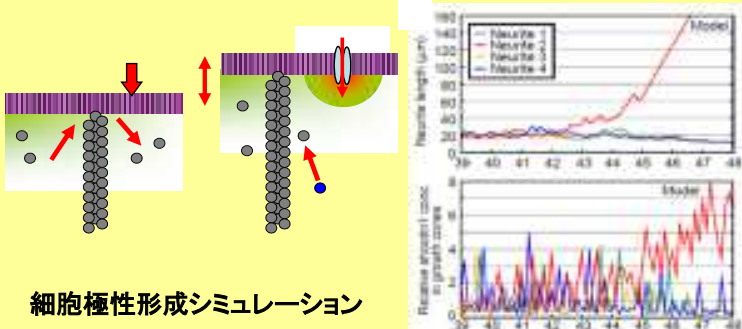


**開発目標:** 脳全体レベルでの入出力関係を再現できるモデルとシミュレーションソフトウェアを開発、次世代スーパーコンピュータの性能を引き出して、脳の柔軟な環境適応能力である脳の発達と学習の再現を目指す。

**開発計画:** 神経細胞と局所回路のシミュレータ、および、脳のシミュレータとして昆虫の匂い情報処理回路モデル、ほ乳類の網膜モデルとを開発、次世代スーパーコンピュータに適したプログラムとすることで、昆虫の嗅覚系情報処理では実時間で、ほ乳類の視覚系情報処理は60倍の時間で計算可能とする。

## 神経マルチフィジクスシミュレータ

## 局所回路シミュレータNEST



**進捗状況:** 本チームは2008年10月に発足、昨年度は神経マルチフィジクスシミュレータでは神経細胞の極性形成について、分子レベルからのシミュレーションができた。また、局所回路シミュレータNESTでは $10^5$ 個の神経細胞、 $10^9$ 個のシナプスからなる大規模局所回路のシミュレーションができた。並列性能に関しては32000並列までのテストを終了

**2011.1.23時点** <sup>11</sup>

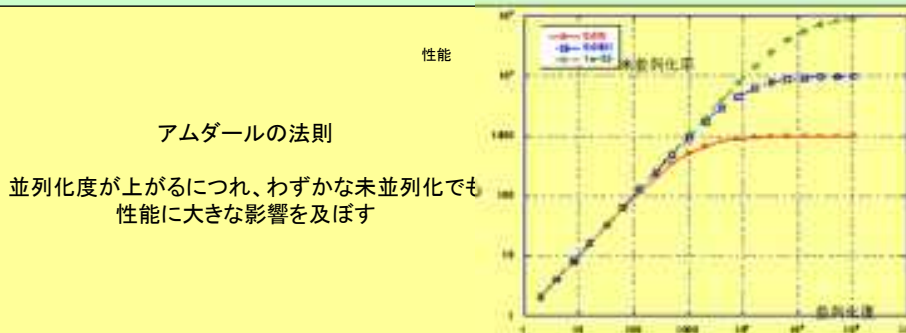


# 生命体基盤ソフトウェア開発・高度化



**開発目標:** プロジェクトで開発される各種ソフトウェアの高性能化、特に次世代スーパーコンピュータ向けの最適化を行い、次世代スーパーコンピュータの生命科学への応用を促進すると同時にグランドチャレンジを実現する。

**開発計画:** コアソフトウェア・基盤ライブラリの開発と他チームアプリケーションの調査・高速化に向けた検討を進め、次世代スーパーコンピュータ実機でのアプリケーションの高速化、チューニングを行う。



**進捗状況:** 本チームは2007年11月に発足、これまでに分子動力学計算高速コアソフトウェア (cppmd) の解析・開発、量子化学計算ソフトウェアの高性能化を行うと同時に、基盤ライブラリ・可視化ソフトウェアの調査・開発を進めた。Cppmdでは8000並列、可視化ソフトでも100並列での動作テストを行った。

**2011.1.23時点** <sup>12</sup>

	ソフトウェア名称	説明
分子(8)	Platypus-MM/CG	マルチコピー・マルチスケール分子シミュレーション法開発の基盤となるクラスライブラリ
	Platypus-REIN	レプリカ交換分子動力学計算インターフェイス
	MARBLE	全原子分子動力学計算
	CafeMol	粗視化モデル計算
	ProteinDF	密度汎関数法に基づくタンパク質全電子波動関数計算
	Platypus-QM/MM-FE	ハイブリッドQM/MM反応自由エネルギー計算
	Platypus-QM	量子化学計算
	Platypus-QM/MM	量子化学計算/分子動力学計算
細胞	RICS	細胞シミュレーションプラットフォーム
臓器全身(4)	ZZ-EFSI	全身ボクセルシミュレーション(ボクセル構造流体連成解析プログラム)
	ZZ-DOSE	重粒子線治療シミュレーション
	HIFU	低侵襲治療シミュレーション(ボクセル超音波伝播プログラム)
	UTHeart	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

	ソフトウェア名称	説明
データ解析融合(9)	ParaHaplo	ハプロタイプ関連解析に於ける統計検定を行うためのソフトウェア
	NGS analyzer	次世代シーケンス解析プログラム
	ExRAT	拡張RAT法による2SNP組合せの全ゲノム関連解析ソフトウェア
	SiGN-BN (SiGN)	大規模遺伝子制御ネットワーク推定プログラム
	SiGN-L1 (L1GN)	再帰的正則化法による生体内分子の大規模ネットワーク推定プログラム
	SiGN-SSM (SSM)	状態空間モデルによる時系列データからの遺伝子ネットワーク推定プログラム
	SBiP (※)	データ解析融合プラットフォーム
	LiSDAS	生命体データ同化プログラム
	MEGADOCK	網羅的タンパク質ドッキング解析プログラム
脳神経(5)	NEST	Neural Simulation Tool
	GMDN	Cortical Microcircuit Developed on NEST
	VSM	全視覚系モデルによる視覚情報処理の解析(視覚系シミュレーションのための共有プラットフォーム)
	NeuroMorphoKit	神経細胞形態シミュレーションキット
	IOSSIM	昆虫嗅覚系全脳シミュレータ
基盤(4)	cppmd	大規模並列MDコアプログラム
	LSV (※)	分散並列大規模データ可視化システム
	SPHERE (※)	アプリケーションモデルウェア
	VLSVL (※)	大規模仮想化合物ライブラリ

(※)は、フロンティアや可視化など直接「京」で稼働させることを前提としていないソフト

## ISLiMのアプリ開発状況

Phase I < 256

Phase II < 8192

Phase III on K

II-0

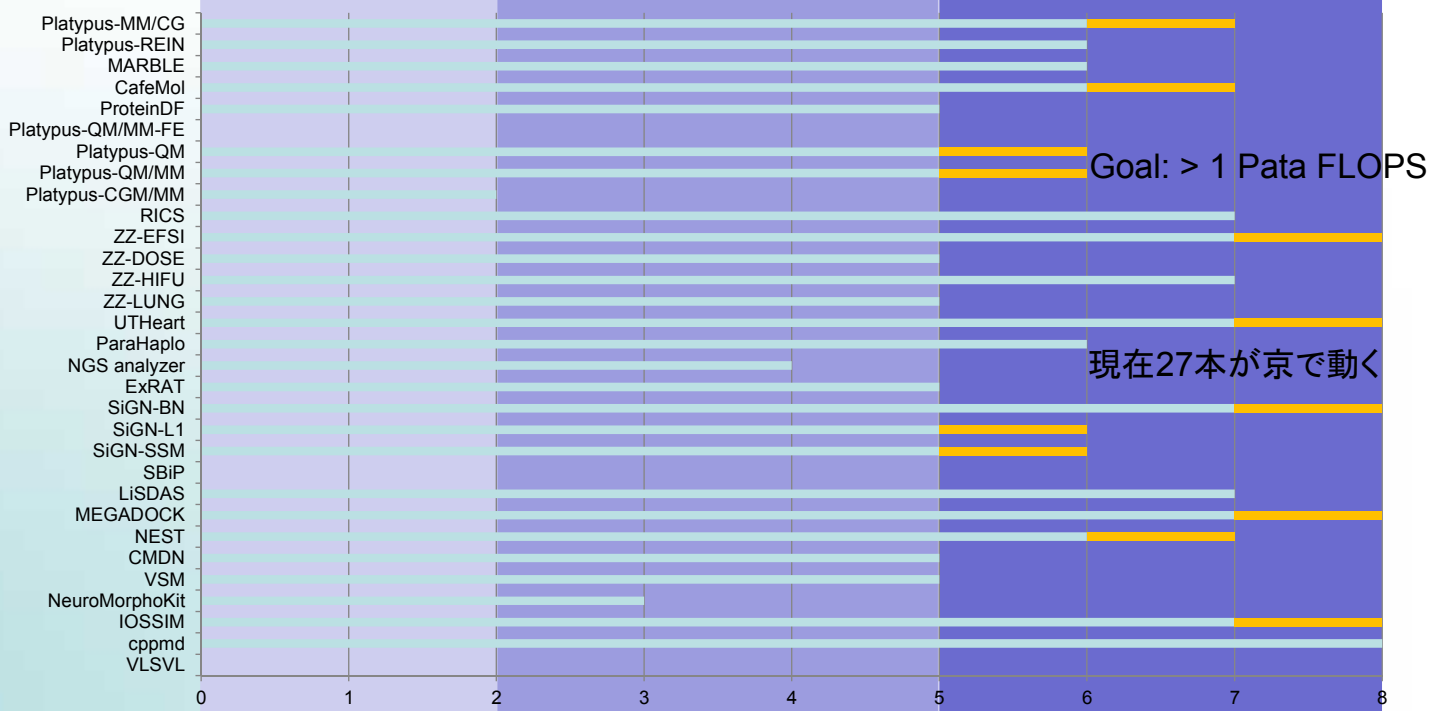
II-1

II-2

1024nds

12288nds

82944nds



cppmd, ZZ-EFSI, UT-Heartの三つは2Peta FLOPSを超えた



## 第一走者

cppmd(分子動力学)  
ProteinDF  
CafeMol(ヒストンDNA)  
UTHeart  
ZZ-EFSI  
ParaHaplo  
NEST(神経回路網)

完成後すぐに成果  
をアピール



## 第二走者

多剤排出トランスポータ  
MARBLE  
Platypus  
CafeMol  
RICS:肝細胞  
HIFU ZZ-DOSE重イオン粒子  
SIGN SSM L1GNE  
LiSDAS  
MEGADOCK  
CMDN:大脳皮質局所回路モ

完成後、1年以内に  
成果をアピール



## 第三走者

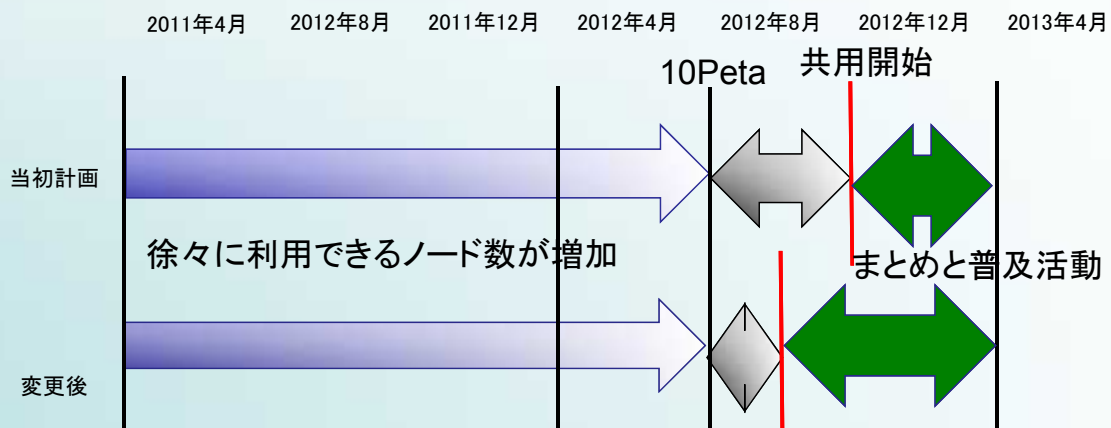
神経系統合シミュレータ  
細胞から臓器へ  
分子から細胞へ

## 中長期目標

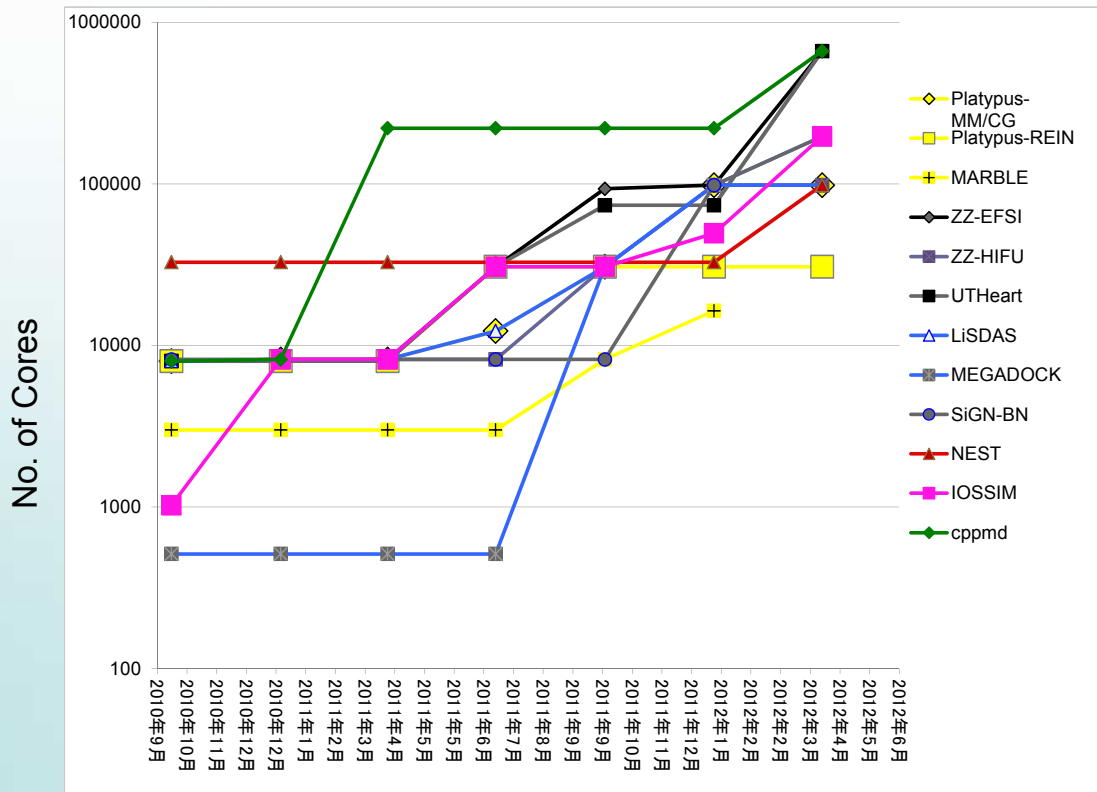
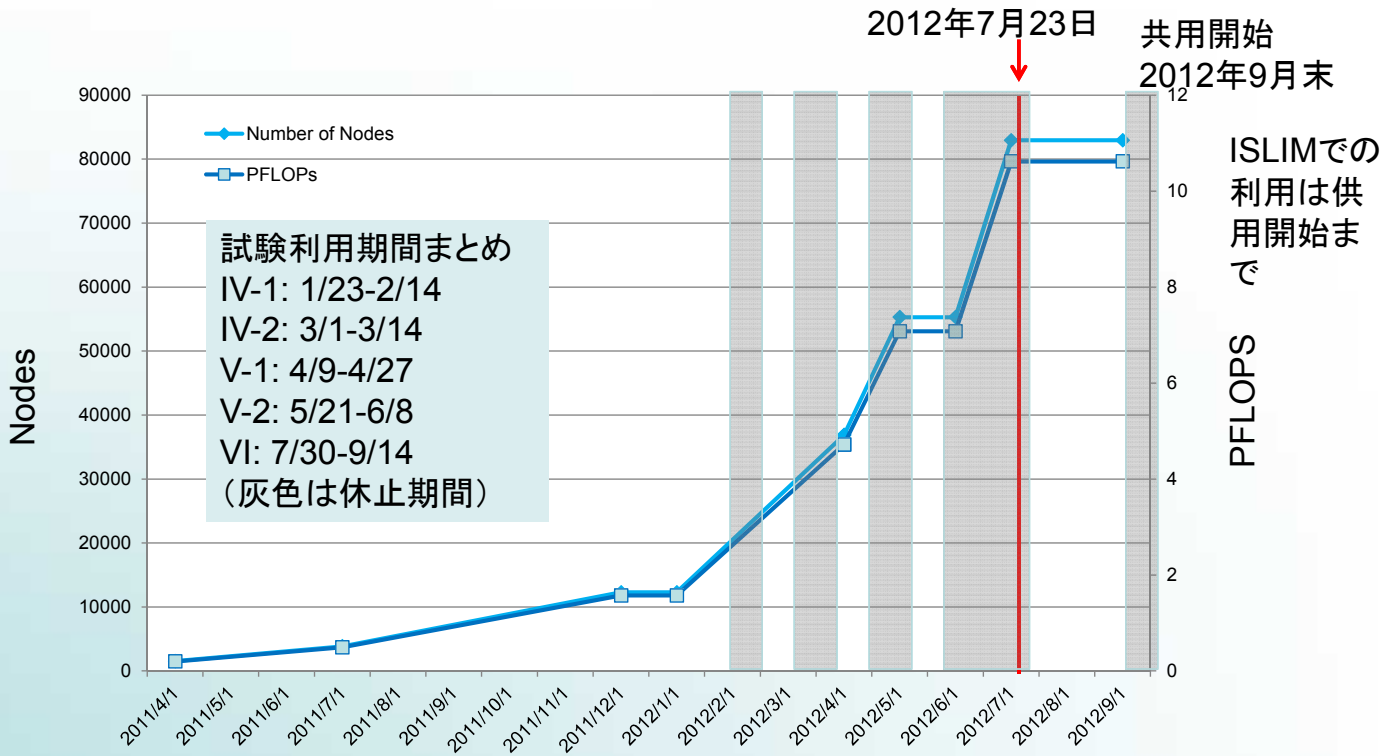
スケールを越えた  
統合的理解に向け  
た取り組み

# 昨年度からの大きな変化

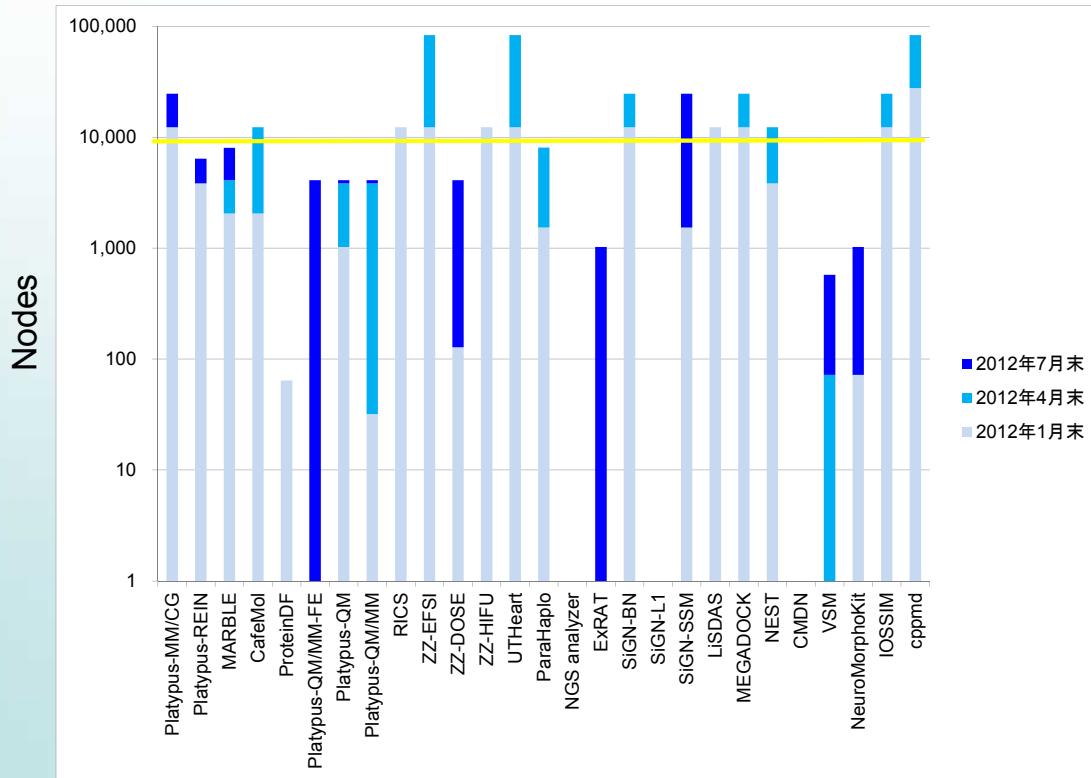
- 平成23年4月より「京」を試験利用。段階的に利用を拡大、ソフトをチューニング（「京」共用開始前までが本プロジェクトで利用できる期間）
- 当初計画では、平成24年11月末まで利用可能  
【「京」完成から約4ヶ月間】
- 共用開始を平成24年9月末に変更（平成24年2月7日 第8回HPCI計画推進委員会）  
【実質約2ヶ月短縮】



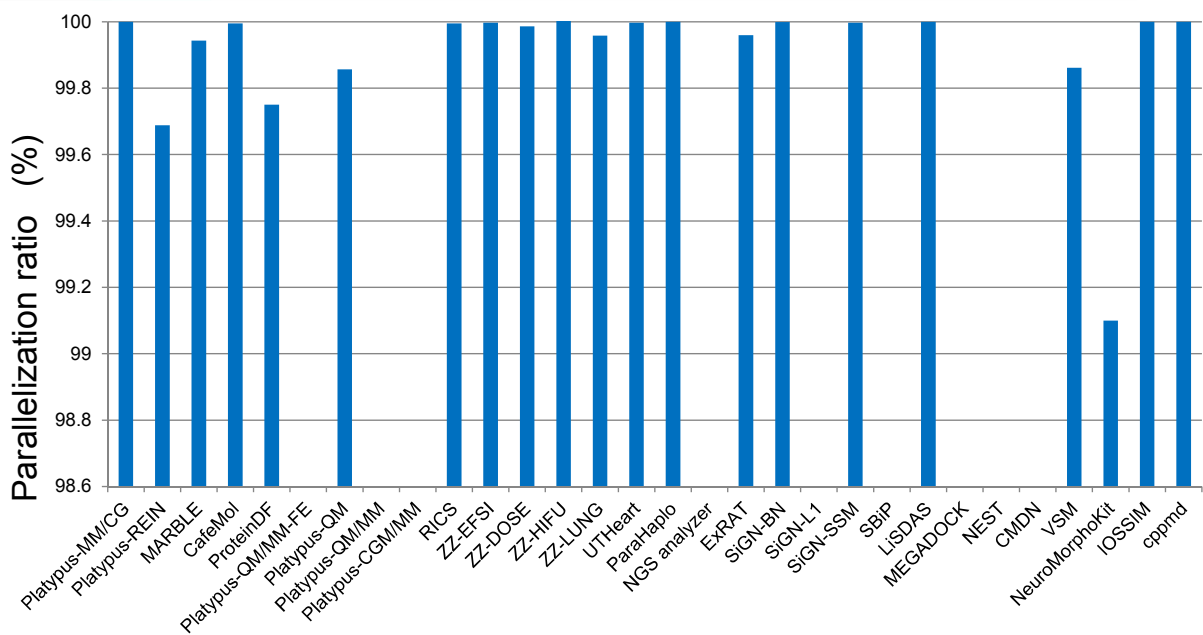




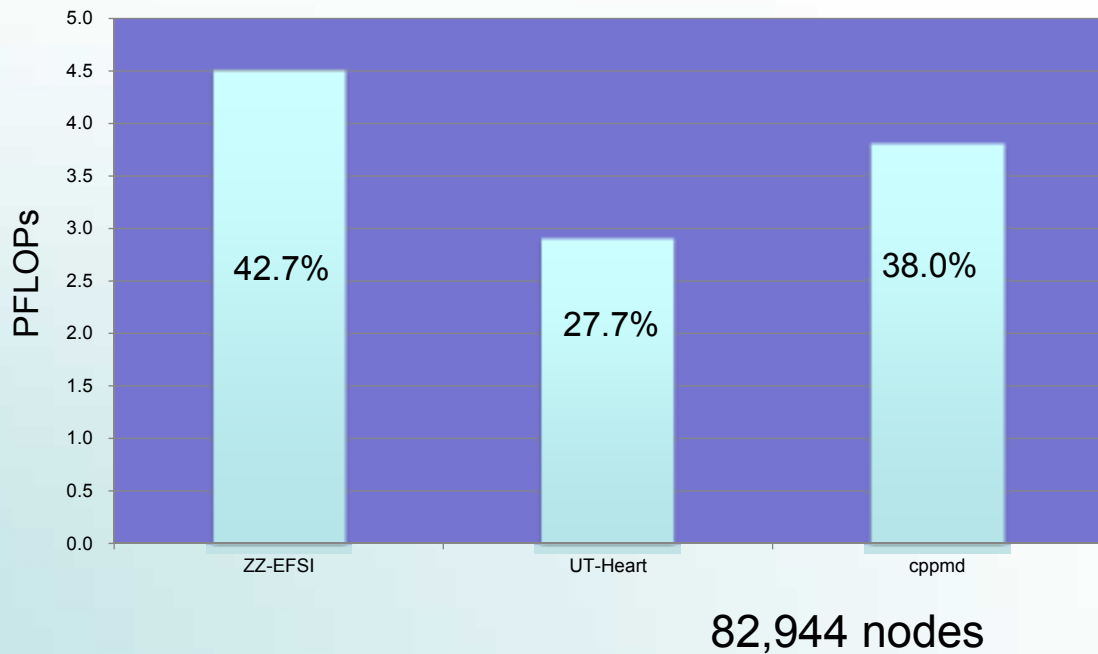
## 全アプリの京で使用した最大ノード数 27本中13本が10,000ノード以上を達成



## 並列化率



## 京全システムを使った性能測定結果(3本)



21

## まとめ

- ソフトウェア開発は当初計画通りで、順調
  - 京の1/8以上を使っても性能が落ちないソフトを13本(全体の約1/3)開発
  - 京のソフトとして最高の実効性能を出したZZ-EFSIを初め、cppmd、UT-Heart、ZZ-HIFUで高い実行効率を達成
  - 現在マニュアルを順次整備中、既に多くのソフトがダウンロード可能
- ただし、京の供用開始前倒しで
  - 京の利用が2ヶ月短縮(京全体が使える期間が4か月から2ヶ月に半減)
  - この影響で当初予定していた京のフルシステムを使った計算は、実質UT-Heartだけ
- 京での計算の結果は徐々にまとまりつつあり、成果が出てきている
- 来年度以降もこのプロジェクトで開発したソフトウェアは利用可能なように整備を続ける予定

22